On the Existence of a Certain Class of Nonlinear Stochastic Processes

by

Dialid Santiago Ramírez

THESIS

Submitted to the University of Warwick

for the degree of

Doctor of Philosophy

Department of Statistics

December 2015

THE UNIVERSITY OF WARWICK

Dialid Santiago Ramírez
On the Existence of a Certain Class of Nonlinear Stochastic Processes
The University of Warwick
Department of Statistics
Coventry, UK
CV4 7AL

Contents

To all those brave women around the world
who have dared to pursue their dreams.

" *La poesía y la matemática son los dos polos extremos del lenguaje. Más allá de ellos no hay nada — el territorio de lo indecible; entre ellos, el territorio inmenso pero finito, de la conversación.*

— **Octavio Paz**
Corriente Alterna (1990)

" *Archimedes will be remembered when Aeschylus is forgotten, because languages die and mathematical ideas do not. "Immortality" may be a silly word, but probably a mathematician has the best chance of whatever it may mean.*

— **G. H. Hardy**
A Mathematician's Apology (1941)

Acknowledgements

I would like to thank my supervisors Sigurd Assing and Vassili Kolokoltsov, for giving me advice, guidance, and support when I most needed, for answering my questions, and clarifying my confused mind, for your infinite patience, and most importantly for teaching me not just about mathematics but about life.

I am also very grateful to Jean-Francois for allowing me to visit the Mathematics Institute of the University of Chile in November 2014. Thanks for his hospitality during this visit, and for many useful discussions.

I am hugely grateful to my beloved partner Javier, for helping me to get through one of the hardest periods of my life. For his unfailing love, support, encouragement and understanding. This would not have been possible without you. You made me strong.

I also want to thank my family: my parents Dialid Ramírez and Isaias Santiago, my little sister Astrid, and last but not least my grandfather José Santiago. A mis padres y a mi hermana por su confianza y por su amor incondicional. A mi padre y a mi abuelo, por transmitirme su genuino interés en las matemáticas y enseñarme que la creatividad no conoce fronteras.

My thanks go to the many friends I have made here at Warwick over the course of my PhD, for making this such an enjoyable time. To my friends: Yang, Silvia, and Pier for making the office the perfect place not just to write this thesis but to feel relaxed, comfortable and happy.

Finally, I acknowledge funding from the Mexican Council of Science and Technology and from the Statistics Department from The University of Warwick.

Declaration

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for any degree. The work presented was carried out by the author except where explicitly indicated otherwise by references.

Abstract

In this thesis, we investigate a class of stochastic processes whose definition can be achieved by formulating a nonlinear martingale problem and subsequently proving its well-posedness. This class includes so-called nonlinear Markov processes, such as McKean-Vlasov processes and nonlinear diffusions, but also non-Markovian versions of those. Roughly speaking, these processes are characterised by the fact that the evolution of their realisations depends on a particular finite dimensional distribution of the process itself. To formalise our idea we need to specify three components: (1) a collection of delay points which determines the finite dimensional distributions to be considered in the nonlinearity; (2) a family of operators which describes the evolution of the marginal probability distributions of the process; and (3) an initial condition which characterises the process on an initial period of time defined by the collection of delay points. Given these three elements we are able to formulate rigorously a nonlinear martingale problem and investigate its well-posedness.

Our main results, which can be found in Chapter 4, provide sufficient conditions to guarantee the existence of a unique solution to the nonlinear martingale problem. The proof consists of three parts: constructing an approximating sequence of "standard" stochastic processes – together with a sequence of related curves of probability measures – proving its convergence, and finally demonstrating that its limit satisfies the martingale problem. To accomplish the proof we require a decomposition akin to the one provided by Ito's formula. The reason why the classical Ito's formula cannot be applied is that we need a decomposition for functions depending on the process at a finite number of non-anticipating times and not just on the process at the current time. To overcome this difficulty we establish an appropriate Ito-type formula by using Skorohod integration theory. The material related to this formula can be found in Chapter 3.

In addition, in Chapter 5 we prove the existence of solutions of a class of nonlinear SDEs with unbounded coefficients by using a different approach which was proposed in Kolokoltsov, 2010 and allows to investigate a class of nonlinear stochastic processes. Finally, we present two examples of nonlinear SDEs in Chapter 6. The purpose of such examples is twofold, first illustrate that the conditions for existence of solutions are sufficient but not necessary; and second to show potential applications. The idea is to propose stochastic volatility models with nonlinear dependence. In particular, we set two models via SDEs.

Nomenclature

$\mathbb{R}$	Set of real numbers
$(a, b]$	Set of all real numbers x such that $a < x \leqslant b$
$\mathbb{R}^+$	Set of positive real numbers
$\mathbb{R}^d$	Set of d-dimensional real vectors
$\|x\|$	Euclidean norm of the real vector $x \in \mathbb{R}^d$
x^T	Transpose of a vector (or matrix) x
$\mathrm{tr}(M)$	Trace of the matrix M
A^c	Complement of the set A
$1_A(\cdot)$	Indicator function of the set A
$\log(\cdot)$	Natural logarithm function
$f'(x)$	First derivative of a function $f : \mathbb{R} \mapsto \mathbb{R}$
$f''(x)$	Second derivative of a function $f : \mathbb{R} \mapsto \mathbb{R}$
$f^{(k)}(x)$	Derivative of order k of a function $f : \mathbb{R} \mapsto \mathbb{R}$
$D_x^{(k)} f(x)$	Derivative of order k of a function $f : \mathbb{R}^d \mapsto \mathbb{R}$
$\dfrac{\partial f}{\partial t}$	Partial derivative of f with respect to t
$\mathbb{P}(E)$	Probability of the event E
$\mathbb{E}[X]$	Expectation of a random variable X
$\mathrm{Var}[X]$	Variance of a random variable X
$\mathrm{Cov}[X, Y]$	Covariance between the random variables X and Y.
$\mathcal{L}(X)$	Distribution of a random variable X
$N(\mu, \sigma^2)$	Gaussian (normal) distribution with mean μ and variance σ^2
$N(\mu, \sum)$	Multivariate Gaussian (normal) distribution with mean vector μ and covariance matrix $\sum$

1 Introduction

> *Most mathematical questions suggested by nature are genuinely nonlinear... The study of such questions is still, after two or three hundred years, in its infancy. Only a few on the simplest examples are understood in any really satisfactory way. I believe this direction will be a principal theme in the future.*
>
> — **Henry P. McKean**
>
> Some Mathematical Coincidences (2003)

THIS thesis is concerned with the analytic question about the existence and uniqueness of stochastic processes. More precisely, we investigate a class of stochastic processes characterised by families of second order differential operators whose coefficients depend on probability distributions defined on Euclidean spaces. Intuitively, we can describe these processes by the fact that their present state depends on at least one non-anticipating finite dimensional probability distribution of the process itself. This class includes McKean-Vlasov processes and nonlinear diffusions, both characterised by their dependence on the distribution of the current state of the process, but also processes with more complicated dependence. In order to state the main goal of the thesis let us introduce a family of integro-differential operators

$$A[\mu] : C^2(\mathbb{R}) \to C(\mathbb{R}), \qquad \mu \in \mathcal{P}(\mathbb{R}^{m+1}),$$

given by

$$A[\mu]f(x) = \frac{1}{2}\sigma(x,\mu)\frac{d^2}{dx^2}f(x) + b(x,\mu)\frac{d}{dx}f(x). \tag{1.1}$$

Here, the functions $\sigma : \mathbb{R} \times \mathcal{P}(\mathbb{R}^{m+1}) \to \mathbb{R}$ and $b : \mathbb{R} \times \mathcal{P}(\mathbb{R}^{m+1}) \to \mathbb{R}$ are measurable and continuous with respect to the first componend for every $\mu \in \mathcal{P}(\mathbb{R}^{m+1})$. It is worth noticing that if we take $m = 0$, then the family of operators given in (1.1) corresponds to a diffusion process with non-linear[1] coefficients. Also, if we take $m = 0$ and consider coefficients depending only on the probability distribution, i.e., $\sigma(\mu), b(\mu)$, then we are in the case of the so-called non-linear Levy processes (see Kolokoltsov, 2012 for details about this kind of processes). Nevertheless, it is not completely clear that we can identify a stochastic process associated with the family of operators given by equation (1.1) in general. Thus, our aim is to provide an answer to the following questions: Is there a stochastic process corresponding to the family of operators given by equation (1.1)? If so, is such a process uniquely determined? In this work we will restrict ourselves to the case when $d = 1$ for simplicity. In order to formalise the above questions we introduce a nonlinear martingale problem. Roughly speaking, this problem involves finding a stochastic process X such that, for nice test functions f, the process

$$f(X_t) - f(X_s) - \int_s^t A[\mu(X_{\leqslant u})]f(X_u)\,du, \qquad t \geqslant s,$$

is a martingale with respect to the filtration generated by X. Here, the notation $\mu(X_{\leqslant u})$ means that the probability measure depends on the path of the process up to the time u. In fact, we are going to consider $\mu(X_{\leqslant u})$ to be a non-anticipating, finite dimensional — $(m+1)$ dimensional to be precise — measure[2] of X. By using the martingale formulation we find sufficient conditions to guarantee existence of a stochastic process associated with the family of operators given in (1.1).

[1] Non-linear in the sense that the coefficients depend on distribution of the process itself.
[2] Notice that this is not a random measure.

In the next section, we present a review of some remarkable works on nonlinear stochastic processes. This review provides a framework and allows the reader to see how our results fit into the big mathematical picture.

1.1 A Brief Survey of the Theory

The connection between the study of Partial Differential Equations (PDEs) and probability theory has shown to be very fruitful for both areas. In particular, it was the key to introduce the so-called nonlinear stochastic processes characterised by their correspondence with nonlinear parabolic differential equations. The study of this class of processes was initiated in the late sixties by H. P. McKean with the publication of his seminal papers 'A Class of Markov Processes Associated with Nonlinear Parabolic Equations' (McKean, 1966) and 'Propagation of chaos for a class of nonlinear parabolic equations' (McKean, 1967). He studied the following class of nonlinear partial differential equations

$$\begin{aligned}\frac{\partial u}{\partial t}(t,x) = \frac{1}{2}\sum_{i,j=1}^{d} \frac{\partial^2 u}{\partial x_i \partial x_j}(t,x)\left\{a_{ij}(x,u(t,\cdot))u(t,x)\right\} \\ -\frac{1}{2}\sum_{i=1}^{d}\frac{\partial u}{\partial x_i}(t,x)\{b_i(x,u(t,\cdot))u(t,x)\}, \qquad t>0, x\in\mathbb{R}^d,\end{aligned} \tag{1.2}$$

where $a_{ij} : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ and $b : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$. Equations of this form arise in fluid mechanics and they include the Boltzmann equation, Landau equation, the granular media equation and many others. McKean observed that the well-known relationship between parabolic equations and Markov processes could be generalised to nonlinear parabolic equations of the form (1.2) giving place to a new class of stochastic processes. Nowadays, such processes are called McKean-Vlasov processes.

The work of McKean opened a new avenue of research from both theoretical and applied points of view. These processes have been studied extensively by several authors and under different sets of conditions. A good introduction to this topic can be found in Sznitman (1991). Further progress on relaxing the assumptions to prove existence of these processes was done by Funaki (1985); Benachour et al. (1998); Benachour, Roynette, and Vallois (1998) and Benedetto, Caglioti, and Pulvirenti (1997).

It is worth mentioning that McKean-Vlasov equations represent just a subset of the class of SDEs whose coefficients depend on the probability distribution of the solution. One can think of more general situations, e.g. coefficients depending on the whole family of time marginal distributions of the process, or on those before the current time, etc. In general, processes which have these kinds of dependence are called nonlinear stochastic processes in the sense of McKean. Hereafter, we will refer to them just as nonlinear stochastic processes for simplicity. A great introduction to more general nonlinear Markov processes can be found in Kolokoltsov (2010).

1.2 Overview

This work is divided into six chapters, outlined as follows.

The first two chapters constitute an introduction for the thesis. In Chapter 1, a detailed overview of the thesis is given. Motivation and main ideas of the thesis are provided.

In Chapter 2, background material and general notation are given. Basic definitions, terminology and some preliminary classical results are provided in this chapter as well.

In Chapter 3, we introduce one-dimensional processes of the form

$$Z_t = f(X_{t-\pi_m}, \cdots X_{t-\pi_1}, X_t),$$

where X is a one dimensional Itô diffusion, $0 = \pi_0 \leqslant \pi_1 \leqslant \cdots \leqslant \pi_m = \pi$ are real numbers, and $f : \mathbb{R}^{m+1} \to \mathbb{R}$ is a smooth function. A Itô type formula for such processes is established.

In Chapter 4, the nonlinear martingale problem is formalised as follows: First, consider a family of operators $\{A[\mu] : C^2(\mathbb{R}) \to C(\mathbb{R})\}$, with common domain $\mathcal{D}$ containing the set $C_c^2(\mathbb{R})$, given by

$$A[\mu]f(x) = \frac{1}{2}\sigma^2(x,\mu)\frac{d^2}{dx^2}f(x) + b(x,\mu)\frac{d}{dx}f(x),$$

where $\sigma : \mathbb{R} \times \mathcal{P}(\mathbb{R}^{m+1}) \to \mathbb{R}$, and $b : \mathbb{R} \times \mathcal{P}(\mathbb{R}^{m+1}) \to \mathbb{R}$ will be called the diffusion and the drift coefficient, respectively; $\Pi = \{0 = \pi_0 \leqslant \pi_1 \leqslant \cdots \leqslant \pi_m = \pi\}$ a collection of delay points, and $\{X_t^0, 0 \leqslant t \leqslant \pi\}$ an stochastic process which will play the role of initial condition. A stochastic process $X = \{X_t, t \geqslant 0\}$ defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is said to be a solution of the nonlinear martingale problem for the triplet $(A[\mu], \Pi, X^0)$ if and only if the process

$$M_t^f = f(X_t) - f(X_\pi) - \int_\pi^t A[\mathcal{L}(X_u, X_{u-\pi_1}, \cdots, X_{u-\pi})]f(X_u)du, \qquad t \geqslant \pi,$$

is an $\mathcal{F}^X$-martingale for each $f \in \mathcal{D}$, and $\mathbb{P} \circ X_t^{-1} = \mathbb{P} \circ (X_t^0)^{-1}$ for every $t \in [0, \pi]$. Our main results provide sufficient conditions on the coefficients to guarantee the existence and uniqueness of a solution for this type of nonlinear martingale problem.

In Chapter 5 we focus on stochastic processes described by nonlinear SDEs of the form

$$\begin{cases} dX_t = b(X_t, \mu_t)dt + \sqrt{2}dW_t \quad \text{where} \quad \mu_t = \mathcal{L}(X_t), \ t \geqslant 0, \\ X_0 \ \text{ given}, \end{cases} \tag{1.3}$$

where $b(X, \mu)$ is of the form

$$b(x, \mu) = \int \beta(x, u)\mu(du),$$

and $\beta : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ is continuous but only locally bounded. Our main result provides sufficient conditions on β for the existence of a weak solution of equation (1.3).

Finally, in Chapter 6, we discuss two examples of nonlinear SDEs. The purpose of this chapter is to illustrate the fact that the class of nonlinear processes that we investigate can be useful in the context of applications.

First, we consider a stock price modelled by a geometric Brownian motion with random volatility coefficient. That is,

$$\begin{cases} dS_t = rS_t dt + \sqrt{Y_t(\mathcal{L}(S_{\leqslant t}))}S_t dW_t, \quad t > 0; \\ S_0 \ \text{ given}, \end{cases} \tag{1.4}$$

where r is a positive constant corresponding to the interest rate, $W = \{W_t : t \geqslant 0\}$ is a standard Brownian motion, and $Y_t = Y_t(\mathcal{L}(S_{\leqslant t}))$ is a stochastic quantity, representing the volatility of the stock price, depending on the distribution of the process S up to the time t. To begin with we study the case where

$$Y_t = \begin{cases} g(t), & 0 \leqslant t < \tau, \\ \frac{1}{\tau}\mathbb{E}\left[\left(\log \frac{S_t}{S_{t-\tau}} - \mathbb{E}\left[\log \frac{S_t}{S_{t-\tau}}\right]\right)^2\right], & t \geqslant \tau, \end{cases} \tag{1.5}$$

for some fixed constant $\tau > 0$ and given $g(t)$. We prove that in this case equation (1.4) can be reduced to the following ordinary differential equation of retarded type:

$$y'(t) - \frac{1}{\tau}y(t) + \frac{1}{\tau}y(t-\tau) = 0, \qquad t \geqslant \tau,$$

with initial condition $y(t) = g(t)$, for every $0 \leqslant t < \tau$. Using this relationship we prove existence and uniqueness of the solution and investigate its long-time behaviour.

Besides, we set a more involved case where Y_t is a stochastic process which takes two values y_1, y_2 (with $y_1 < y_2$) and whose dynamics can heuristically be described as follows: $Y_t = y_1$, for $0 \leqslant t \leqslant \tau$, and for $t \geqslant \tau$ the process jumps from y_1 to y_2 with rate

$$\lambda_t = \mathbb{E}[f^-(S_t - S_{t-\tau})], \tag{1.6}$$

where

$$f^-(s) = \begin{cases} -f(s) & \text{if } f(s) < 0, \\ 0 & \text{otherwise,} \end{cases}$$

for some smooth function f, and from y_2 to y_1 with constant rate $\lambda > 0$. We believe that the existence and uniqueness of a solution $(S_t, Y_t, t \geqslant 0)$ can be proved by following the ideas presented in Chapter 4 but left this as a future research project.

2 Background Material

> *Mathematicians find the same sort of beauty others find in enchanting pieces of music...The beauty in mathematical structures, however, cannot be appreciated without understanding of a group of numerical formulae that express laws of logic. Only mathematicians can read "musical scores" containing many numerical formulae, and play that "music" in their hearts.*
>
> — **Kiyosi Itô**
>
> My Sixty Years in Studies of Probability Theory (1998)

The purpose of this chapter is twofold, to establish general notation and to provide the reader with a theoretical framework which includes basic definitions and results that will be required in the subsequent chapters. In the first section, we present some general notation. The second section is devoted to basic definitions from the theory of stochastic processes. The main properties of Brownian motion as well as various constructions are given in the third section. In addition, we provide some general notions of Itô Calculus for Brownian motion. This construction is only a particular case of Stochastic Calculus for semi-martingales and can be found in the literature (see e.g. Karatzas and Shreve, 1991 or Kolokoltsov, 2011). However, we choose to present it here for two reasons: our results do not require the general framework and moreover this general framework is abstract enough, so that a short introduction in a simple case may be useful as an introduction for the next chapter. Finally, the last part focuses on the Malliavin Calculus and the Skorohod integral.

2.1 General Notation

In this section, we introduce general notation and definitions that will be used throughout the thesis.

Operators. Let X be a Banach space equipped with the norm $\|\cdot\|_X$, then X^* will denote its dual or conjugated space equipped with the dual norm, i.e.:

$$\|x^*\|_{X^*} = \sup\{x^*x : \|x\|_X \leqslant 1\}, \qquad \forall x^* \in X^*.$$

Let $L : X \to Y$ be a linear operator between Banach spaces, then the norm of L is given by

$$\|L\|_{\mathcal{L}(X,Y)} = \sup\{\|Lx\|_Y : \|x\|_X \leqslant 1\}.$$

L^p Spaces. Let (S, Σ, μ) be a measure space and $1 \leqslant p < \infty$. The space $L^p(S, \Sigma, \mu)$ consist of equivalence classes of real valued measurable functions $f : S \to \mathbb{R}$ which satisfy

$$\int_0^\infty |f|^p \, d\mu < \infty,$$

where two measurable functions are equivalent if they are equal μ-almost surely. Then if f belongs to $L^p(S, \Sigma, \mu)$ we define the L^p norm of f by

$$\|f\|_{L^p(S,\Sigma,\mu)} := \left(\int_0^\infty |f|^p \, d\mu\right)^{1/p},$$

or simply

$$\|f\|_{L^p} := \left(\int_0^\infty |f|^p \, d\mu\right)^{1/p} < \infty,$$

In particular, for $S = [0, \infty)$. $\Sigma = \mathcal{B}([0, \infty))$ and μ being the Lebesgue measure, we will simply use $L^p([0, \infty))$ to denote $L^p(S, \Sigma, \mu)$.

General Function Spaces. Let $C(\mathbb{R}^d)$ denote the family of continuous real valued functions defined on $\mathbb{R}^d$. Let $C_b(\mathbb{R}^d)$ denote the space of all bounded continuous functions on $\mathbb{R}^d$ equipped with the supremum norm

$$\|f\| = \sup_{x \in \mathbb{R}^d} |f(x)|.$$

Let $C_0(\mathbb{R}^d)$ denote the space of continuous functions with compact support. Let $C_\infty(\mathbb{R}^d)$ denote the space of continuous functions which vanish at infinity. That is, all the continuous functions f such that for every $\epsilon > 0$ there exists a compact set $K = K(\epsilon)$ such that

$$|f(x)| < \epsilon, \qquad \text{for every} \quad x \in K^c.$$

Differentiable Function Spaces. For any integer $k > 0$, let $C^k(\mathbb{R}^d)$ denote the class of continuous functions $f : \mathbb{R}^d \to \mathbb{R}$ with continuous derivatives up to order k, and $C^\infty(\mathbb{R}^d)$ will denote the class of continuous functions with continuous derivatives of all orders.

Let $C_b^\infty(\mathbb{R}^d)$ denote the set of continuous functions $f : \mathbb{R}^d \to \mathbb{R}$ such that f and all its partial derivatives are bounded. Finally, $C_0^\infty(\mathbb{R}^d)$ denotes the set of infinitely differentiable functions $f : \mathbb{R}^d \to \mathbb{R}$ with compact support, while $C_p^\infty(\mathbb{R}^d)$ denotes the set of infinitely continuously differentiable functions $f : \mathbb{R}^d \to \mathbb{R}$ such that f and all of its partial derivatives have polynomial growth.

Weighted Spaces. The introduction of a weight makes it possible to enlarge or shrink ordinary function spaces and hence make them more useful for certain problems. The following spaces will play this role for us. For $q > 0$, let $C_q(\mathbb{R}^d)$ denote the space of continuous functions f such that

$$\sup_{x \in \mathbb{R}^d} \left\{ \frac{|f(x)|}{1 + |x|^q} \right\} < \infty.$$

Let $C_{q,\infty}(\mathbb{R}^d)$ denote the family of continuous functions $f : \mathbb{R}^d \to \mathbb{R}$ such that ${f(x)}/{1+\|x\|^q}$ belongs to $C_\infty(\mathbb{R}^d)$. Clearly this family is contained in $C_q(\mathbb{R}^d)$. Besides, $C_{q,\infty}(\mathbb{R}^d)$ equipped with the norm

$$|f|_q = \sup_{x\in\mathbb{R}^d}\left\{\frac{|f(x)|}{1+|x|^q}\right\},$$

is a Banach space (see e.g. Summers, 1970). Similarly, for q, r and s positive numbers, let $C^2_{q,r,s}(\mathbb{R}^d)$ denote the space of twice differentiable continuous functions $f : \mathbb{R}^d \to \mathbb{R}$ such that

$$\sup_{x\in\mathbb{R}^d}\left\{\frac{|f(x)|}{1+|x|^q} + \sum_{i=1}^{d}\frac{\left|\frac{\partial f(x)}{\partial x_i}\right|}{1+|x|^r} + \sum_{i,j=1}^{d}\frac{\left|\frac{\partial^2 f(x)}{\partial x_i \partial x_j}\right|}{1+|x|^s}\right\} < \infty.$$

Finally, let $C^2_{q,r,s,\infty}(\mathbb{R}^d)$ denote the space of twice differentiable continuous functions f, such that

$$\frac{f(x)}{1+|x|^q}, \qquad \frac{{\partial f(x)}/{\partial x_i}}{1+|x|^r}, \quad 1 \leqslant i \leqslant d, \quad \text{and} \quad \frac{{\partial^2 f(x)}/{\partial x_i \partial x_j}}{1+|x|^s}, \quad 1 \leqslant i, j \leqslant d,$$

belong to $C_\infty(\mathbb{R}^d)$, which when equipped with the above norm is a Banach space (see e.g. Summers, 1970).

Probability measures. Let $\mathcal{P}(\mathbb{R}^d)$ denote the family of probability measures defined on $\mathbb{R}^d$. For $1 \leqslant q < \infty$ we will use $\mathcal{P}_q$ to denote the family of all probability measures $\mu \in \mathcal{P}(\mathbb{R}^d)$ such that

$$\int_{\mathbb{R}^d}(1+|u|^q)\mu(du) < \infty.$$

For $1 \leqslant q < \infty$ and $N > 0$, we will use $\mathcal{P}^N_q$ to denote the family of all probability measures $\mu \in \mathcal{P}(\mathbb{R}^d)$ such that

$$\int_{\mathbb{R}^d}(1+|u|^q)\mu(du) < N.$$

Through our exposition we will work with sequence of elements of these families. More precisely, we will establish convergence of such sequences. Thus, we will need to specify the metric to define distance and convergence concepts. One option is to use the Wasserstein-Kantorovich distance. For any μ_1 and μ_2 in $\mathcal{P}_q$, the Wasserstein-Kantorovich metric W_p between them is given by

$$W_p(\mu_1, \mu_2) = \inf \left(\int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^p \gamma(dx, dy) \right)^{1/p},$$

where the infimum is taken over all the probability measures defined on $\mathbb{R}^d \times \mathbb{R}^d$ whose marginals coincide with μ_1 and μ_2. Another option that one can use is to explode the fact that these families are embedded in a certain normed vector space $\mathcal{M}$ and then just consider the norm inherited by such embedding. That is, for any μ_1 and μ_2 in $\mathcal{P}_q \subset \mathcal{M}$, the distance between them is given by

$$d(\mu_1, \mu_2) = |\mu_1 - \mu_2|_{\mathcal{M}},$$

where $|\cdot|_{\mathcal{M}}$ denotes the norm induced by the embedding of $\mathcal{P}_q$ in $\mathcal{M}$.

Curves of probability measures. Let $\mathcal{Q}$ be a subset of $\mathcal{P}(\mathbb{R}^d)$. For $\xi \in \mathcal{Q}$ fixed and $T > 0$, let $C_\xi([0,T], \mathcal{Q})$ be the collection of continuous curves of probability measures given by

$$\nu_\cdot = \{\nu_t : \nu_t \in \mathcal{Q} \ \text{ for all } \ 0 < t \leqslant T \ \text{ and } \ \nu_0 = \xi\}.$$

If $\mathcal{Q}$ is contained in some normed vector space $\mathcal{M}$, then the set $C_\xi([0,T]; \mathcal{Q})$ can be equipped with the norm

$$\|\nu\|_{\mathcal{M}} := \sup_{0 \leqslant t \leqslant T} |\nu_t|_{\mathcal{M}},$$

where $|\cdot|_{\mathcal{M}}$ denotes the norm induced by the embedding of $\mathcal{Q}$ in $\mathcal{M}$.

Further concepts and notation will be introduced in subsequent sections as and where appropriate.

2.2 Stochastic Processes

To begin with, let us recall some general definitions about stochastic processes.

Definition 2.1. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $(S, \mathcal{B}, \mu)$ a measure space. An S-valued **stochastic process** is a collection of random variables, defined on $(\Omega, \mathcal{F}, \mathbb{P})$ and taking values on S, indexed by a totally ordered set Θ. That is, a stochastic process X is a collection of the form $\{X_\theta : \theta \in \Theta\}$ where each X_θ is a random variable taking values on S.

The space Ω is called the **sample space** while the space S is called the **state space** of the process. In this work, we are going to focus on stochastic processes taking values in the d-dimensional Euclidean space $\mathbb{R}^d$. That is, we are going to consider measure spaces of the form $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda)$ where $\mathcal{B}(\mathbb{R}^d)$ denotes the Borel σ-algebra of $\mathbb{R}^d$, and λ the Lebesgue measure defined on $\mathbb{R}^d$.

The set Θ is called the **index set**. We are going to assume that $\Theta = [0, T]$ for some $T > 0$ or $T = \infty$. This temporal feature of a stochastic process suggests a flow of time, in which at every given time $t \in [0, T]$ we have a notion of the past, the present and the future.

Definition 2.2. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A **forward filtration**, or simply **filtration**, is a sequence of sub-σ-algebras $\{\mathcal{F}_t\}_{t \geqslant 0}$ such that $\mathcal{F}_t \subseteq \mathcal{F}$ for each $t \geqslant 0$, and $\mathcal{F}_{t_1} \subseteq \mathcal{F}_{t_2}$ whenever $t_1 \leqslant t_2$.

Similarly, a **backward filtration** is a sequence of σ-algebras $\{\mathcal{F}_t\}_{t \geqslant 0}$ with $\mathcal{F}_t \subseteq \mathcal{F}$ for each $t \geqslant 0$ and such that $\mathcal{F}_{t_2} \subseteq \mathcal{F}_{t_1}$ whenever $t_1 \leqslant t_2$.

Definition 2.3. A probability space $(\Omega, \mathcal{F}, \mathbb{P})$ endowed with a filtration $\{\mathcal{F}_t\}_{t \geqslant 0}$ is called a **filtered probability space** or **stochastic basis**.

Definition 2.4. Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geqslant 0}, , \mathbb{P})$ be a filtered probability space.

(i) The filtration $\{\mathcal{F}_t\}_{t\geqslant 0}$ is said to be **complete** if $(\Omega, \mathcal{F}, \mathbb{P})$ is complete and if $\mathcal{F}_0$ contains all the $\mathbb{P}$-null sets.

(ii) The filtration $\{\mathcal{F}_t\}_{t\geqslant 0}$ is said to satisfy the **usual hypothesis** if it is complete and right continuos, that is $\mathcal{F}_t = \mathcal{F}_{t+}$, where

$$\mathcal{F}_{t+} = \bigcap_{u>t} \mathcal{F}_u.$$

Definition 2.5. Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geqslant 0},, \mathbb{P})$ be a filtered probability space satisfying the usual hypothesis and $X = \{X_t : t \geqslant 0\}$ an $\mathbb{R}^d$-valued stochastic process defined on this space. We say that X is **adapted** to the filtration $\mathcal{F}$ if $X_t \in \mathcal{F}_t$ for all $t \geqslant 0$.

Definition 2.6. For each $t \in [0, T]$, we will denote by $\mathcal{F}_t^X$ the σ-algebra generated by the collection of random variables $\{X_s : 0 \leqslant s \leqslant t\}$ and the null sets in $\mathcal{F}$. This family $\{\mathcal{F}_t^X\}$ will be called **natural filtration of the process** X.

Clearly, any stochastic process X is adapted to its natural filtration.

Definition 2.7. Let $X = \{X_t : t \geqslant 0\}$ be an $\mathbb{R}^d$-valued stochastic process defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We say that X is **almost surely continuous** (right continuous, left continuous) if for almost every $\omega \in \Omega$, the mapping $t \mapsto X_t(\omega)$ is continuous (right continuous, left continuous).

Now, let us consider two $\mathbb{R}^d$ valued stochastic processes $X = \{X_t : t \geqslant 0\}$ and $\tilde{X} = \{\tilde{X}_t : t \geqslant 0\}$, defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$. It is clear that X and $\tilde{X}$ are equal, regarded as functions in (ω, t), if and only if

$$X_t(\omega) = \tilde{X}_t(\omega), \qquad \text{for all } \ (\omega, t) \in \Omega \times [0, \infty).$$

However, in probability theory there are weaker notions of stochastic processes being equal.

Definition 2.8. The process $\tilde{X}$ is said to be a **modification** of X, if

$$\mathbb{P}(X_t = \tilde{X}_t) = 1, \qquad \text{for every} \quad t \geqslant 0.$$

Definition 2.9. The processes X and $\tilde{X}$ have **the same finite-dimensional distributions**, if

$$\mathbb{P}((X_{t_1}, \cdots, X_{t_n}) \in A) = \mathbb{P}((\tilde{X}_{t_1}, \cdots, \tilde{X}_{t_n}) \in A),$$

for any integer $n \geqslant 1$, real numbers $0 \leqslant t_1 < t_2 < \cdots < t_n < \infty$ and $A \in \mathcal{B}(\mathbb{R}^{d \times n})$.

Definition 2.10. The processes X and $\tilde{X}$ are called **indistinguishable**, if

$$\mathbb{P}(X_t = \tilde{X}_t; \forall t \geqslant 0) = 1.$$

Observe that the last definition implies the first one, which in turn yields the second one. But notice that the second definition can be extended to processes defined on different probability spaces while the other two cannot. Such an extension is presented below.

Definition 2.11. Let X and $\tilde{X}$ be stochastic processes defined on $(\Omega, \mathcal{F}, \mathbb{P})$ and $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$, respectively. Then we say that they have **the same finite-dimensional distributions**, if

$$\mathbb{P}((X_{t_1}, \cdots, X_{t_n}) \in A) = \tilde{\mathbb{P}}((\tilde{X}_{t_1}, \cdots, \tilde{X}_{t_n}) \in A),$$

for any integer $n \geqslant 1$, real numbers $0 \leqslant t_1 < t_2 < \cdots < t_n < \infty$ and $A \in \mathcal{B}(\mathbb{R}^{d \times n})$.

2.2.1 Quadratic Variation and Cross Variation

Definition 2.12. The **quadratic variation** of a stochastic process $X = \{X_t : t \geqslant 0\}$ on the time interval $[0, u]$ is defined as

$$[X]_u = \lim \sum_{i=1}^{N} (X_{t_i} - X_{t_{i-1}})^2,$$

whenever the sum converges in probability, where the limit is taken over all partitions P^N of $[0, u]$ with $|P^N| = \max_i |t_i - t_{i-1}| \to 0$.

Definition 2.13. The **crossed variation** between two stochastic process $X = \{X_t : t \geqslant 0\}$ and $Y = \{Y_t : t \geqslant 0\}$ on the time interval $[0, u]$ is defined as

$$[X, Y]_u = \lim \sum_{i=1}^{N} (X_{t_i} - X_{t_{i-1}})(Y_{t_i} - Y_{t_{i-1}}),$$

whenever the sum converges in probability, where the limit is taken over all partition P^N of $[0, u]$ with $|P^N| = \max_i |t_i - t_{i-1}| \to 0$.

2.3 Brownian Motion and Stochastic Calculus

Brownian motion, named after Robert Brown a Scottish botanist, is without doubt one of the best known continuous stochastic processes in probability theory. In 1827, while looking through a microscope at particles trapped inside pollen grains in water, he observed that the particles moved erratically through the water but was not able to determine the mechanisms that caused this kind of motion. In 1905 Albert Einstein published a paper (Einstein, 1905) that explained how the motion observed by Brown was the result of the pollen being moved by the individual water molecules. This explanation of Brownian motion provided a confirmation that atoms and molecules actually existed, and was further verified experimentally by Jean Perrin in 1908. The first mathematical construction is due to Norbert Wiener in 1923.

2.3.1 Brownian Motion

As we mentioned, Brownian motion or Wiener process is one of the most important stochastic processes and it is the basis for modelling in many areas of stochastic analysis. Its formal definition is as follows.

Definition 2.14. A real valued stochastic process $W = \{W_t : t \geqslant 0\}$ is called a **linear Brownian motion or simply a Brownian motion** started at $x \in \mathbb{R}$, if the following holds:

1. $W_0 = x$.

2. Independence of increments: For all times $0 \leqslant t_1 \leqslant \cdots \leqslant t_{n-1} \leqslant t_n$, the increments $W_{t_n} - W_{t_{n-1}}, W_{t_{n-1}} - W_{t_{n-2}}, \cdots, W_{t_2} - W_{t_1}$ are independent random variables.

3. Normal increments: For all $t \geqslant 0$ and $\epsilon > 0$, the increments $W_{t+\epsilon} - W_t$ are normally distributed with expectation zero and variance ϵ.

4. Continuity of paths: The mapping $t \mapsto W_t$ is continuous almost surely.

   If $x = 0$, then we say that $\{W_t : t \geqslant 0\}$ is a **standard Brownian motion**.

This definition has been extended to the multidimensional framework as follows.

Definition 2.15. Let d be an integer and $W^1, \cdots, W^d$ independent linear Brownian motions starting at $x_1, x_2, \cdots, x_d \in \mathbb{R}$, respectively. The stochastic process $W = \{W_t : t \geqslant 0\}$ given by

$$W_t = (W_t^1, W_t^2, \cdots, W_t^d)^\intercal, \qquad t \geqslant 0,$$

is called a **d-dimensional Brownian motion** starting at $(x_1, x_2, \cdots, x_d)^\intercal \in \mathbb{R}^d$.

A d-dimensional Brownian motion started at the origin is called a **d-dimensional standard Brownian motion**.

Next, we present stochastic integration with respect to the Brownian motion.

2.3.2 Stochastic Integration

First, let us consider a standard Brownian motion W on an interval $[0, T] \subset \mathbb{R}^+$, that can also be infinite, defined on the canonical probability space $(\Omega, \mathcal{F}, \mathbb{P})$. That is, Ω is the space of real continuous functions defined on $[0, T]$, denoted by $C([0, T], \mathbb{R})$, and $\mathbb{P}$ is a probability measure defined on the Borel σ-algebra $\mathcal{B}(\Omega)$ such that the canonical process $W_t(\omega) = \omega(t)$ is a zero mean Gaussian process with covariance $\mathrm{Cov}[W_s, W_t] = \min\{s, t\}$. The σ-algebra $\mathcal{F}$ will be the completion of $\mathcal{B}(\Omega)$ with respect to the measure $\mathbb{P}$.

For each $t \in [0, T]$, we will denote by $\mathcal{F}_t^W$ the σ-algebra generated by the random variables $\{W_s : 0 \leqslant s \leqslant t\}$ and the null sets in $\mathcal{F}$. Then $\{\mathcal{F}_t^W\}$ is a forward filtration[1], i.e., $\mathcal{F}_s^W \subseteq \mathcal{F}_t^W$ whenever $0 \leqslant s \leqslant t \leqslant T$.

Definition 2.16. A stochastic process $\{X_t : 0 \leqslant t \leqslant T\}$ will be called **adapted to the forward filtration of Brownian motion** if X_t is $\mathcal{F}_t^W$ measurable for all $t \in [0, T]$.

Similarly, for each $t \in [0, T]$, we will denote by $\mathcal{F}_W^t$ the σ-field generated by the random variables $\{W_T - W_s : t \leqslant s \leqslant T\}$ and the null sets of $\mathcal{F}$. Then $\{\mathcal{F}_W^t\}$ is a backward filtration, i.e., $\mathcal{F}_W^s \supseteq \mathcal{F}_W^t$ whenever $0 \leqslant s \leqslant t \leqslant T$.

Definition 2.17. A stochastic process $\{Y_t : 0 \leqslant t \leqslant T\}$ will be called **adapted to the backward filtration of Brownian motion** if Y_t is $\mathcal{F}_W^t$-measurable for all $t \in [0, T]$.

[1] In fact, this is the natural filtration of W. See Definition 2.6.

Now, let us recall the definitions of **forward and backward stochastic integrals**. Let $(P^N)_{N \geqslant 1}$ be a sequence of partitions

$$P^N = \{0 = s_1^N < \cdots < s_N^N = T\}, \qquad N \geqslant 1,$$

such that

$$\lim_{N \to \infty} |P^N| = \lim_{N \to \infty} \max_i |s_i^N - s_{i-1}^N| = 0.$$

Hereafter, we will write s_i instead of s_i^N for notational convenience.

Definition 2.18. Let $\{X_t : 0 \leqslant t \leqslant T\}$ be a real valued continuous stochastic process adapted to the forward filtration $\mathcal{F}_t^W$, and $\Phi \in C(\mathbb{R})$. Then, the forward Itô integral, or simply **Itô integral**, of $\Phi(X_t)$ **with respect to the Brownian motion** can be defined by

$$\int_0^T \Phi(X_s) dW_s = \lim_{N \to \infty} \sum_{i=1}^N \Phi(X_{s_i})(W_{s_{i+1}} - W_{s_i}),$$

where the right hand side refers to the limit in probability.

Definition 2.19. Let $\{Y_t : 0 \leqslant t \leqslant T\}$ be a real valued continuous process adapted to the backward filtration $\mathcal{F}_W^t$, and $\Psi \in C(\mathbb{R})$. Then, the **backwards Itô integral** of $\Psi(Y_t)$ **with respect to the Brownian motion** can be defined by

$$\int_0^T \Psi(Y_s) \overleftarrow{dW_s} = \lim_{N \to \infty} \sum_{i=1}^N \Psi(Y_{s_{i+1}})(W_{s_{i+1}} - W_{s_i}),$$

where the right hand side refers to the limit in probability.

2.3.3 Itô's Formula

Theorem 2.1. *(see e.g. Itô, 1944, Kunita and Watanabe, 1967) Let* $X = \{X_t : t \geqslant 0\}$ *be a continuous semimartingale defined on some probability space* $(\Omega, \mathcal{F}, \mathbb{P})$*, and* $f \in C^2(\mathbb{R})$*. Then*

$$f(X_t) = f(X_0) + \int_0^t f'(X_s)dX_s + \frac{1}{2}\int_0^t f''(X_t)d[X]_t, \qquad t \geqslant 0, \text{a.s.}$$

2.4 Skorohod Integral and Malliavin Calculus

In this section, we introduce some results from the theory of Malliavin calculus. Our intention is to provide the reader with all the concepts and theorems that we require in the proofs of our own results. Many theorems are presented without proofs, or with only an outline of the proof. For further explanations we refer to the excellent monograph by Nualart (2013) that contains all the missing proofs and much more material.

2.4.1 The Malliavin Derivative

First, we recall the definition of the Malliavin derivative which plays the role of a gradient on the Wiener space. In our context the divergence operator can be interpreted as the stochastic integral introduced in Skorokhod (1976).

Definition 2.20. An **isonormal Gaussian process** consists of the following three elements:

1. A real and separable Hilbert space H.

2. A complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

3. A Gaussian process indexed by H, $W = \{W(h) : h \in H\}$, that is, W is a centred Gaussian family of random variables such that $\mathbb{E}[W(h)W(g)] = \langle h, g\rangle_H$.

Remark 2.1. Observe that given such space H, we can always construct $(\Omega, \mathcal{F}, \mathbb{P})$ and W satisfying the above conditions by using Kolmogorov's theorem.

We are going to restrict ourselves to the following particular case.

Definition 2.21. Let $W = \{W_t : 0 \leqslant t \leqslant T\}$ be a d-dimensional standard Brownian motion defined on its canonical probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then the **Isonormal Gaussian process associated with** W is given by the following three elements:

1. Choose $H = L^2([0, T])$.

2. Choose the canonical probability space of W.

3. For each $h \in H$, define $W(h)$ to be the Itô integral

$$W(h) = \int_0^T h(s)dW_s.$$

Definition 2.22. Let $\mathcal{S}_p$ denote the **class of smooth functionals or smooth random variables** of the form

$$F = f(W(h_1), \cdots, W(h_n)), \tag{2.1}$$

where f belongs to $C_p^\infty(\mathbb{R}^n)$, $h_1, \cdots, h_n$ belong to H, and $n \geqslant 1$.

Similarly, we will use $\mathcal{S}_b$ and $\mathcal{S}_0$ to denote the class of smooth random variables F as given by (2.1) but choosing f from $C_b^\infty(\mathbb{R}^n)$ and $C_0^\infty(\mathbb{R}^n)$, respectively. It is clear that $\mathcal{S}_0 \subset \mathcal{S}_b \subset \mathcal{S}_p$.

Remark 2.2. Notice that $\mathcal{S}_0$ is dense in $L^2(\Omega)$.

Definition 2.23. Let $F = f(W(h_1), \cdots, W(h_n))$ be a smooth random variable in $\mathcal{S}_p$. Then the **Malliavin derivative** of F is defined to be the stochastic process $DF = \{D_t F : 0 \leqslant t \leqslant T\}$ given by

$$D_t F = \sum_{i=1}^{n} \partial_i f(W(h_1), \cdots, W(h_n)) h_i(t).$$

Example 2.1. Consider a smooth functional $F = f(W(\pi_1), \cdots, W(\pi_n))$ where $f \in C_p^\infty(\mathbb{R}^n)$ and

$$W(\pi_i) = W(1_{[0,\pi_i]}) = \int_0^{\pi_i} dW_t,$$

for times $\pi_1, \cdots, \pi_n$. Then, its Malliavin derivative is the stochastic process $\{D_t F : t \geqslant 0\}$, given by

$$D_t F = \sum_{i=1}^{n} \partial_i f(W(\pi_1), \cdots, W(\pi_n)) 1_{[0,\pi_i]}(t).$$

Proposition 2.2. *The Malliavin derivative is closable from* $L^2(\Omega)$ *to* $L^2(\Omega; H)$

Proof. This is a particular case of Proposition 1.2.1 in Nualart (2013). ☐

In what follows, we are going to denote the closed extension of the Malliavin derivative operator $\bar{D}$ also by D. Besides, we are going to use $\mathbb{D}^{1,2}$ to denote the **domain of the Malliavin derivative** in $L^2(\Omega)$. That is, $\mathbb{D}^{1,2}$ is the closure of $\mathcal{S}_p$ with respect to the norm

$$\|F\|_{1,2} = \left(\mathbb{E}[|F|^2] + \mathbb{E}[\|DF\|_H^2]\right)^{1/2}.$$

Notice that $\mathbb{D}^{1,2}$ equipped with the scalar product

$$\langle F, G \rangle_{1,2} = \mathbb{E}[FG] + \mathbb{E}[\langle DF, DG \rangle_H],$$

is a Hilbert space.

Remark 2.3. The Malliavin derivative of a random variable F in $\mathbb{D}^{1,2}$ can be seen as a stochastic process $DF = \{D_t F : t \geqslant 0\}$ defined almost surely with respect to the measure

$\lambda \times \mathbb{P}$. This is due to the identification between the Hilbert spaces $L^2(\Omega; H)$ and $L^2([0, \infty) \times \Omega)$.

The next results is the analogous to the chain rule for usual derivatives.

Lemma 2.3. *(Proposition 1.2.3 in in Nualart, 2013) Let $\psi : \mathbb{R}^m \to \mathbb{R}$ be a continuously differentiable function with bounded partial derivatives. Suppose that $F = (F^1, \cdots, F^m)$ is a random vector whose components belong to the space $\mathbb{D}^{1,2}$. Then $\psi(F)$ also belongs to $\mathbb{D}^{1,2}$, and*

$$D(\psi(F)) = \sum_{i=1}^{m} \partial_i \psi(F) DF^i.$$

In the case of Malliavin derivative, the chain rule can be extended to the case of a Lipschitz function as stated in the following result.

Lemma 2.4. *(Proposition 1.2.4 in in Nualart, 2013) Let $\psi : \mathbb{R}^m \to \mathbb{R}$ be a Lipschitz continuous function with Lipschitz constant $\kappa > 0$. Suppose that $F = (F^1, \cdots F^m)$ is a random vector whose components belong to the space $\mathbb{D}^{1,2}$. Then $\psi(F)$ also belongs to $\mathbb{D}^{1,2}$, and there exists a random vector $K = (K^1, \cdots, K^m)$ bounded by κ such that*

$$D(\psi(F)) = \sum_{i=1}^{m} K_i DF^i.$$

Lemma 2.5. *(Lemma 1.2.3 in Nualart, 2013) Let $(F_n)_{n\geqslant 1}$ be a sequence of variables in $\mathbb{D}^{1,2}$ that converges to F in $L^2(\Omega)$ and such that*

$$\sup_{n\geqslant 1} \mathbb{E}[\|DF_n\|_H^2] < \infty.$$

Then, F belongs to $\mathbb{D}^{1,2}$ and the sequence of derivatives $(DF_n)_{n\geqslant 1}$ converges to DF in the weak topology of $L^2(\Omega; H)$.

2.4.2 The Divergence Operator

In this section we introduce the divergence operator which is the adjoint of the closed extension of the Malliavin derivative operator in $L^2(\Omega)$. For our choice of H this operator is also called Skorohod integral.

Definition 2.24. The **divergence operator** $\delta : L^2(\Omega; H) \to L^2(\Omega)$ is the unbounded operator determined by the following characteristics:

1. Its domain $\mathrm{Dom}(\delta)$ is the set of H-valued square integrable random variables $u \in L^2(\Omega; H)$ such that

$$|\mathbb{E}[\langle DF, u\rangle_H]| \leqslant c\|F\|_{L^2(\Omega)},$$

for all $F \in \mathbb{D}^{1,2}$, where c is some constant depending on u.

2. For $u \in \mathrm{Dom}(\delta)$, $\delta(u)$ is the unique element in $L^2(\Omega)$ characterised by

$$\mathbb{E}[F\delta(u)] = \mathbb{E}[\langle DF, u\rangle_H],$$

for any $F \in \mathbb{D}^{1,2}$.

Definition 2.25. The **class of smooth elementary processes** is the class of processes of the form

$$u = \sum_{i=1}^{n} F_i h_i,$$

where F_i belongs to $\mathcal{S}_p$ (the class of smooth random variables), h_i belong to H, for all $i = 1, \cdots, n$. The class of smooth elementary processes will be denoted by $\mathcal{S}_p \otimes H$.

Thus, an arbitrary element of $\mathcal{S}_p \otimes H$ is of the form

$$u = \sum_{i=1}^{n} F_i h_i = \sum_{i=1}^{n} f_i(W(\tilde{h}_{1_i}), \cdots, W(\tilde{h}_{n_i}))h_i, \tag{2.2}$$

where f_i belongs to $C_p^\infty(\mathbb{R}^{n_i})$, $\tilde{h}_{1_i}, \cdots, \tilde{h}_{n_i}$ belong to H, $n_i \geqslant 1$, and h_i belong to H for all $i = 1, \cdots, n$.

Some properties of δ are summarised below.

1. If u belongs to $\mathrm{Dom}(\delta)$, then $\mathbb{E}[\delta(u)] = 0$.
2. δ is a linear and closed operator – as the adjoint of a densely defined unbounded operator.
3. The class of smooth elementary processes $\mathcal{S}_p \otimes H$ is contained in $\mathrm{Dom}(\delta)$. Moreover, for $u \in \mathcal{S}_p \otimes H$, as given in (2.2), we have

$$\delta(u) = \sum_{i=1}^{n} F_i W(h_i) - \sum_{i=1}^{n} \langle DF_i, h_i \rangle_H, \tag{2.3}$$

almost surely.

Next, we present some results which allow us to apply the divergence operator to more general types of random variables.

Proposition 2.6. *(c.f. Prop 1.3.1 in Nualart, 2013) The space $\mathbb{D}^{1,2}(H)$ is included in Dom(δ) and δ is continuous from $\mathbb{D}^{1,2}(H)$ into $L^2(\Omega)$.*

The following propositions allow us to calculate the divergence of products.

Proposition 2.7. *(c.f. Prop 1.3.2 in Nualart, 2013) Let $F \in \mathbb{D}^{1,2}$ and $u \in$Dom(δ). If Fu belongs to $L^2(\Omega; H)$, then Fu belongs to Dom(δ) and*

$$\delta(Fu) = F\delta(u) - \langle DF, u \rangle_H \qquad \text{a.s.},$$

provided that the right hand side of this equation is square integrable.

Finally, the following results provide useful criteria to determine the existence of a divergence.

Proposition 2.8. *(c.f. Prop 1.3.6 in Nualart, 2013) Let* u *be an element of* $L^2(\Omega; H)$ *such that there exists a sequence* $\{u_n\}_{n \geqslant 1}$ *of elements in Dom*(δ) *which converges to* u *in* $L^2(\Omega; H)$. *If there exists* G *in* $L^2(\Omega)$ *such that*

$$\lim_{n \to \infty} \mathbb{E}[\delta(u_n)F] = \mathbb{E}[GF], \qquad \forall F \in \mathcal{S}_p,$$

then u *belongs to Dom*(δ) *and* $\delta(u) = G$.

Proposition 2.9. *Let* u *be an element of* $L^2([0, \infty) \times \Omega)$. *If* u *is an adapted process, then* u *belongs to Dom*(δ). *Moreover,* $\delta(u)$ *coincides with the (forward) Itô integral with respect to the Brownian motion, i.e.,*

$$\delta(u) = \int_0^\infty u_s dW_s, \qquad \text{a.s.}$$

Proof. The proof is analogous to the one of Proposition 1.3.18 in Nualart (2013). □

Finally, if H is $L^2([0, T])$ or $L^2([0, \infty))$ and $u \in \text{Dom}(\delta)$, then $\delta(u)$ is also called **Skorohod integral**. Hereafter, we are going to use either $\delta(u)$ or $\int_0^\infty u_t \delta W_t$, to denote a Skorohod integral. Notice that the Skorohod integral is an extension of the Itô integral in the sense of Proposition 2.9.

2.4.3 Localisation

Consider Malliavin derivate D and divergence δ as discussed in the previous, two sections, and recall that D and δ can be defined on $\mathbb{D}^{1,2}$ and $\mathbb{D}^{1,2}(H) \subseteq \mathrm{Dom}(\delta)$, respectively.

Following the ideas presented in Section 1.3.5 of Nualart (2013), we define $\mathbb{D}^{1,2}_{loc}$ to be the set of random variables F such that there exists a sequence $(\Omega_n, F_n)_{n\geqslant 1} \subseteq \mathcal{F} \times \mathbb{D}^{1,2}$ satisfying:

(i) $\Omega_n \uparrow \Omega$ as $n \to \infty$ a.s.;

(ii) $F = F_n$ a.s. on $\Omega_n, n \geqslant 1$.

Similarly, we define $\mathbb{D}^{1,2}_{loc}(H)$.

Then, D and δ can be extended to $\mathbb{D}^{1,2}_{loc}$ and $\mathbb{D}^{1,2}_{loc}(H)$ by setting $DF = DF_n$ and $\delta(u) = \delta(u_n)$ on $\Omega_n, n \geqslant 1$, where $(\Omega_n, F_n)_{n\geqslant 1}$ and $(\Omega_n, u_n)_{n\geqslant 1}$ are localising sequences of $F \in \mathbb{D}^{1,2}_{loc}$ and $u \in \mathbb{D}^{1,2}_{loc}(H)$, respectively.

Note that $\mathrm{Dom}(\delta)$ itself cannot be localised properly (see Nualart (2013), Section 1.3.5).

3

Itô Type Formula for Delay Vectors

> *It is perfectly true, as the philosophers say, that life must be understood backwards. But they forget the other proposition, that it must be lived forwards.*
>
> — **Søren Kierkegaard**
>
> Journals and Papers (1837)

CONSIDER a diffusion process $X = \{X_t : t \geqslant 0\}$ given by $X_t = X_0 + \int_0^t \sigma(s, X_s) dW_s + \int_0^t b(s, X_s) ds$; and a collection of $(m+1)$ delay points given by $\Pi = \{0 = \pi_0 < \pi_1 < \cdots < \pi_m = \pi\}$. This chapter is concerned with the $\mathbb{R}^{(m+1)}-$valued process $\mathbf{X}_\Pi = \{\mathbf{X}_{\Pi_t} : t \geqslant \pi\}$ given by $\mathbf{X}_{\Pi_t} = (X_{t-\pi}, X_{t-\pi_{m-1}}, \cdots, X_{t-\pi_1}, X_t)$. We establish an Itô type formula for processes of the form $\mathbf{Z} = \{\mathbf{Z}_t : t \geqslant \pi\}$ given by $\mathbf{Z}_t = f(\mathbf{X}_{\Pi_t})$, where $f : \mathbb{R}^{m+1} \to \mathbb{R}$ is a sufficiently smooth function. Our formula provides a decomposition of $\mathbf{Z}$ into a sum of a mean zero process, which is in fact a Skorohod integral process of finite variation which is absolutely continuous w.r.t. the Lebesgue measure.

3.1 Introduction

In brief, Itô's formula establishes that given a real valued semimartingale $X = \{X_t : t \geqslant 0\}$ and a sufficiently smooth function $f : \mathbb{R} \to \mathbb{R}$, the process $Z = \{Z_t : t \geqslant 0\}$ given by $Z_t = f(X_t)$ for all $t \geqslant 0$, is also a semimartingale. Moreover, it provides an explicit Bichteler-Dellacherie decomposition of Z into a sum of a stochastic Itô integral (the local martingale) and a Lebesgue integral (the finite variation process). Itô's formula has shown to be an extremely significant tool for theoretical and applied mathematics. For this reason, Itô's formula has been revisited by several authors through the years. Its simplest version, for one dimensional processes, has been extended to cover multidimensional processes, Hilbert-space valued semimartingales (see Métivier, 1982), and infinite dimensional processes (see e.g. Da Prato and Zabczyk, 2014). These extensions allow us to consider X lying in a wider class of processes. On the other hand, extensions to cover a wider class of functions f have been also made. For instance, Föllmer, Protter, and Shiryayev (1995) studied the case of non smooth functions, and Kunita (1997) the case of random functions such as flows of stochastic differential equations.

However, it is worth noticing that in all cases the process Z is defined as a function depending on X_t the current value of the process of interest. Thus, it is natural to wonder about more general scenarios. For instance, we could think of a function of two variables: the current value of the process and its value at a previous time. More generally, one could think of functions depending on previous states of the process or even on the whole path up to the current time. In this direction, we can mention the work of Ahn (1997) who studied the case when the function depends on $X_{t\wedge s}$ rather than on X_t. That is, Ahn obtained an Ito-type formula for the level process. But, a general Itô's formula for functions depending on previous states of the process, the so-called "tame" functions, was given in 2004 by Hu, Mohammed, and Yan (2004). Later, Dupire (2009) proposed a groundbreaking approach to generalise the Itô formula to the functional setting. His idea consisted of introducing a path-wise derivative for non-anticipative functionals on the space of cádlág — right continuous functions with left limits— functions, and describing the variations of the functional in terms of such derivatives. Afterwards, this idea was generalised by Cont and Fournié (2013) to cover to the space of square-integrable martingales.

In this chapter, we review the Itô type formula given by Hu, Mohammed, and Yan in 2004. This formula, is to our knowledge the only one covering delay functionals. The functional setting of Dupire (2009) and Cont and Fournié (2013) does not apply to delays as the corresponding functionals would not be horizontally differentiable. However, the formula established by Hu, Mohammed, and Yan (2004) goes beyond the semimartingale setup of $X = \{X_t : t \geqslant 0\}$, and hence it requires stronger conditions. These conditions, when applied to diffusions X governed by SDEs, would require the diffusion coefficient to be twice differentiable. The good news is that a modification of the proof is possible if X is a diffusion leading to weaker conditions allowing for Lipschitz continuous coefficients. So, in what follows we will introduce notation and discuss the technical difficulties when dealing with delay-processes.

Then, we will prove an Itô type formula when delay-functionals are applied to Brownian motion but our proof will be different to the proof given by Hu, Mohammed, and Yan (2004).

Finally, we weaker the conditions given by Hu, Mohammed, and Yan (2004) in the case of diffusions X governed by SDEs. For this purpose, we benefit from our proof in the case of Brownian motion because we can now refer to those steps in the proof by Hu, Mohammed, and Yan which need modification.

3.1.1 Setting the Problem

The scenario that we are going to investigate is described as follows. Fix a finite time horizon $T > 0$ and consider the following stochastic differential equation

$$\begin{cases} dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t, \\ X_0 \ \text{ given}, \end{cases} \tag{3.1}$$

where $W = \{W_t : 0 \leqslant t \leqslant T\}$ is a standard linear Brownian motion defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$; X_0 is an independent random variable defined on the same probability space, and the functions

$$b : [0, T] \times \mathbb{R} \to \mathbb{R},$$

$$\sigma : [0, T] \times \mathbb{R} \to \mathbb{R},$$

which will be referred to as the **drift** and **diffusion coefficient**, respectively, are measurable functions satisfying:

Assumption 3.1. (Global Lipschitz Condition) There exists a constant $K > 0$ such that

$$|b(t, x) - b(t, y)| + |\sigma(t, x) - \sigma(t, y)| \leqslant K|x - y|,$$

for any $0 \leqslant t \leqslant T$ and $x, y \in \mathbb{R}$.

Assumption 3.2. (Boundedness) There exists a constant $M > 0$ such that

$$\sup_{0 \leqslant t \leqslant T} \{|b(t, 0)| + |\sigma(t, 0)|\} < M.$$

It has been proved (see for example Karatzas and Shreve, 1991; Stroock and Varadhan, 2007; Kolokoltsov, 2010) that these assumptions are sufficient to guarantee the existence of a unique solution to equation (3.1) satisfying

$$X_t = X_0 + \int_0^t b(s, X_s)ds + \int_0^t \sigma(s, X_s)dW_s, \qquad 0 \leqslant t \leqslant T, \qquad a.s.$$

Furthermore, if $\mathbb{E}[|X_0|^p] < \infty$ for some $p \geqslant 2$,

$$\mathbb{E}\left[\sup_{0 \leqslant t \leqslant T} |X_t|^p\right] \leqslant C_0, \tag{3.2}$$

where C_0 is a positive constant depending on p, T, K, M and $\mathbb{E}[|X_0|^p]$.

Next, consider a finite collection of points given by

$$\Pi = \{0 = \pi_0 < \pi_1 < \cdots < \pi_m = \pi\}, \quad \text{for some integer} \quad m > 0,$$

and let us introduce the $\mathbb{R}^{m+1}$-valued process $X_\Pi = \{X_{\Pi_t} : \pi \leqslant t \leqslant T\}$ given by

$$X_{\Pi_t} = (X_{t-\pi_m}, X_{t-\pi_{m-1}}, \cdots, X_{t-\pi_1}, X_t), \qquad \pi \leqslant t \leqslant T. \tag{3.3}$$

Our aim is to prove that, under assumptions 3.1 and 3.2 it is possible to establish an Itô type formula for processes of the form

$$f(X_{\Pi_t}) = f(X_{t-\pi_m}, X_{t-\pi_{m-1}}, \cdots, X_{t-\pi_1}, X_t), \qquad \pi \leqslant t \leqslant T. \tag{3.4}$$

where $f : \mathbb{R}^{m+1} \to \mathbb{R}$ is a twice continuously differentiable function.

The rest of the chapter is structured as follows. In Section 3.2, we introduce some definitions and notation in order to simplify the treatment of random vectors of the form (3.3). In Section 3.3, we investigate processes of the form (3.4) and obtain an Itô type formula in terms of a Skorohod integral and a Lebesgue integral.

3.2 Preliminary Material

The purpose of this section is to introduce some definitions as well as notation which helps us to simplify the treatment of random vectors of the form (3.3).

3.2.1 Back-shifted Processes

To begin with, let us introduce the following definition of a back-shifted process.

Definition 3.1. Let $X = \{X_t : t \geqslant 0\}$ be a stochastic process defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. For $\pi > 0$, the **back-shifted process** $X^\pi = \{X_t^\pi : t \geqslant \pi\}$ is given by $X_t^\pi = X_{t-\pi}$.

Example 3.1. Let $W = \{W_t : t \geqslant 0\}$ be a standard Brownian motion defined on the canonical probability space $(\Omega, \mathcal{F}, \mathbb{P})$. For any $\pi > 0$, we can consider the back-shifted process $W^\pi = \{W_t^\pi : t \geqslant \pi\}$ given by $W_t^\pi = W_{t-\pi}$ for all $t \geqslant \pi$. It is worth noticing the following properties of the back-shifted Brownian motion W^π.

(1) $W_\pi^\pi = 0$, almost surely.

(2) For each $t \geqslant \pi$, W_t^π follows a Gaussian distribution with mean zero and variance equal to $t - \pi$.

(3) If $\{\mathcal{G}_t\}_{t \geqslant \pi}$ is the filtration generated by W^π, then

$$\mathbb{E}[W_t^\pi | \mathcal{G}_s] = \mathbb{E}[W_{t-\pi} | \mathcal{F}_{s-\pi}] = W_{s-\pi} = W_s^\pi, \qquad \text{a.s.},$$

for all $t > s \geqslant \pi$, and hence the back-shifted Brownian motion is a martingale with respect to its natural filtration.

Remark 3.1. The back-shifted Brownian W^π is not a martingale with respect to $\{\mathcal{F}_t^W\}_{t \geqslant \pi}$ the filtration generated by the Brownian motion. The back-shifted process W^π is adapted to $\{\mathcal{F}_t^W\}_{t \geqslant \pi}$ and integrable. Moreover, we have

$$\mathbb{E}[W_t^\pi | \mathcal{F}_s^W] = \mathbb{E}[W_{t-\pi} | \mathcal{F}_s^W] = \begin{cases} W_s & \text{if } t - \pi \geqslant s, \\ W_t^\pi & \text{if } t - \pi < s, \end{cases}$$

almost surely, for all $t > s \geqslant \pi$.

The next results illustrate the fact that the cross variation between a process and its back shifted version vanishes for some processes. This property will be shown to be useful in the next section.

Proposition 3.1. *Let W be a standard Brownian motion and π a positive constant. The crossed variation between W and the back-shifted process W^{π} is zero.*

Proof. First, let us recall the definition of crossed variation. If for any sequence $(P^N)_{N\geqslant 1}$ of partitions of the interval $[\pi, t]$ such that $|P^N| \to 0$, as N tends to infinity, the limit

$$\lim_{N\to\infty}\sum_{i=1}^{N}(W_{s_i^N} - W_{s_{i-1}^N})(W^{\pi}_{s_i^N} - W^{\pi}_{s_{i-1}^N}),$$

exists in probability and is the same, we define such limit to be $[W, W^{\pi}]$. Thus, it is sufficient to consider an arbitrary sequence of partitions

$$P^N = \{\pi = s_0^N < s_1^N < \cdots < s_N^N = t\}, \qquad N \geqslant 1,$$

with $|P^N| \to 0$, and to prove that

$$\mathcal{E}^N = \mathbb{E}\left[\left(\sum_{i=1}^{N}(W_{s_i^N} - W_{s_{i-1}^N})(W^{\pi}_{s_i^N} - W^{\pi}_{s_{i-1}^N})\right)^2\right],$$

converges to zero as N goes to infinity.

By using a simple algebraic decomposition we obtain

$$\mathcal{E}^N = \mathcal{E}_1^N + \mathcal{E}_2^N, \tag{3.5}$$

where

$$\mathcal{E}_1^N = \sum_{i=1}^{N}\mathbb{E}\left[(W_{s_i^N} - W_{s_{i-1}^N})^2(W_{s_i^N-\pi} - W_{s_{i-1}^N-\pi})^2\right],$$

and

$$\mathcal{E}_2^N = 2\sum_{i>j}\mathbb{E}\left[(W_{s_i^N} - W_{s_{i-1}^N})(W_{s_i^N-\pi} - W_{s_{i-1}^N-\pi})(W_{s_j^N} - W_{s_{j-1}^N})(W_{s_j^N-\pi} - W_{s_{j-1}^N-\pi})\right].$$

We will study the convergence of these sums as N goes to infinity.

Hereafter, we will write s_i instead of s_i^N, for notational convenience. Besides, we are going to assume that $|P^N| < \pi$ for all $N \geqslant 1$. We can do this without loss of generality since we are interested only on partitions whose mesh tends to zero as N goes to infinity. This additional assumption implies that

$$s_{i-1} - \pi \leqslant s_i - \pi < s_{i-1} < s_i, \quad \text{for all} \quad i = 1, \cdots, N,$$

and consequently, $W_{s_i} - W_{s_{i-1}}$ and $W_{s_i-\pi} - W_{s_{i-1}-\pi}$ are independent random variables. Therefore

$$\mathbb{E}\left[(W_{s_i} - W_{s_{i-1}})^2 (W_{s_i-\pi} - W_{s_{i-1}-\pi})^2\right] \quad = \quad (s_i - s_{i-1})^2,$$

and then

$$\mathcal{E}_1^N = \sum_{i=1}^{N} (s_i - s_{i-1})^2 \leqslant |P^N|(t - \pi). \tag{3.6}$$

On the other hand, observe that, for all $i > j$, we have

$$s_{j-1} - \pi < s_j - \pi \leqslant s_{i-1} - \pi < s_i - \pi.$$

Besides, notice that

$$s_{i-1} - \pi \leqslant s_i - \pi < s_{i-1} < s_i \quad \text{and} \quad s_{j-1} - \pi \leqslant s_j - \pi < s_{j-1} < s_j,$$

since we are assuming that $|P^N| < \pi$. Hence,

$$s_{j-1} - \pi \leqslant s_j - \pi < s_{j-1} < s_j \leqslant s_{i-1} < s_i,$$

whenever $i > j$. Hence, by using the property of independent increments we can conclude that

$$\begin{aligned}
&\mathbb{E}\left[(W_{s_i} - W_{s_{i-1}})(W_{s_i-\pi} - W_{s_{i-1}-\pi})(W_{s_j} - W_{s_{j-1}})(W_{s_j-\pi} - W_{s_{j-1}-\pi})\right] \\
&= \mathbb{E}\left[(W_{s_i} - W_{s_{i-1}})(W_{s_i-\pi} - W_{s_{i-1}-\pi})(W_{s_j} - W_{s_{j-1}})\right] \mathbb{E}\left[(W_{s_j-\pi} - W_{s_{j-1}-\pi})\right] \\
&= \mathbb{E}\left[(W_{s_i} - W_{s_{i-1}})(W_{s_i-\pi} - W_{s_{i-1}-\pi})(W_{s_j} - W_{s_{j-1}})\right] \cdot 0 \\
&= 0,
\end{aligned}$$

whenever $i > j$, which implies that

$$\mathcal{E}_2^N = 0. \tag{3.7}$$

Substituting (3.6) and (3.7) into (3.5), we obtain

$$\mathcal{E}^N \leqslant |P^N|(\pi - t),$$

and the result follows by taking the limit as N goes to infinity. $\square$

An easy consequence of this result is given as follows.

Corollary 3.2. *Let W be a standard Brownian motion. For any shifts $\pi_1 \neq \pi_2$, the crossed variation between the back-shifted processes W^{π_1} and W^{π_2} is zero.*

Proof. Assume $\pi_2 > \pi_1$ without loss of generality. Then

$$W^{\pi_2} = (W^{\pi_1})^{\pi} \quad \text{with} \quad \pi = \pi_2 - \pi_1 > 0,$$

and the result follows copying the proof of Proposition 3.1. $\square$

3.2.2 Delay Random Vectors

Now, we are ready to introduce the delay vectors. The idea is quite simple and self-explanatory, so let us go straight to the definition.

Definition 3.2. Let $X = \{X_t : t \geqslant 0\}$ be a stochastic process defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. For $\pi \geqslant 0$, and an integer $m \geqslant 0$, let

$$\Pi = \{0 = \pi_0 < \pi_1 < \cdots < \pi_m = \pi\},$$

be a a collection of delay points. The **delay random vector** corresponding to X and Π is defined as the process $\mathbf{X}_\Pi = \{\mathbf{X}_{\Pi_t} : t \geqslant \pi\}$ given by

$$\mathbf{X}_{\Pi_t} = (X_{t-\pi_m}, X_{t-\pi_{m-1}}, \cdots, X_{t-\pi_1}, X_t), \qquad t \geqslant \pi.$$

It is worth noticing that the components of a delay random vector are simply back shifted processes. Thus, delay vectors can be expressed in terms of back-shifted processes as follows:

$$\mathbf{X}_{\Pi_t} = (X_t^{\pi_m}, X_t^{\pi_{m-1}}, \cdots, X_t^{\pi_1}, X_t), \qquad t \geqslant \pi.$$

Remark 3.2. The i-th component of the delay random vector is defined for every $t \geqslant \pi_{m-(i-1)}$ for $i = 1, \cdots, m+1$. However, the delay random vector is defined only for $t \geqslant \pi$ since it is only after this time that all its components are defined.

Example 3.2. Consider a standard Brownian motion $W = \{W_t : t \geqslant 0\}$ and the following collection of delay points $\Pi = \{0 = \pi_0 < \pi_1 < \cdots < \pi_m = \pi\}$. Then the delay random vector $\mathbf{W}_\Pi = \{\mathbf{W}_{\Pi_t} : t \geqslant \pi\}$ is given by

$$\mathbf{W}_{\Pi_t} = (W_{t-\pi}, W_{t-\pi_{m-1}}, \cdots, W_t), \qquad t \geqslant \pi.$$

Clearly, each component of $\boldsymbol{W}_{\Pi_t}$ has a normal distribution. In fact, the independence of the increments

$$W_{t-\pi} - W_{t-\pi_{m-1}}, W_{t-\pi_{m-1}} - W_{t-\pi_{m-2}}, \cdots, W_{t-\pi_1} - W_t,$$

implies that $\boldsymbol{W}_{\Pi_t}$ is jointly normally distributed for each $t \geqslant \pi$. More precisely, $\boldsymbol{W}_{\Pi_t}$ follows a multivariate normal distribution with mean

$$\mu = (0, \cdots, 0),$$

and covariance matrix $\Sigma(t) = (\Sigma_{ik}(t))_{0 \leqslant i,k \leqslant m}$ given by

$$\begin{aligned} \Sigma_{ik}(t) &= \mathrm{Cov}(W_{t-\pi_{m-i}}, W_{t-\pi_{m-k}}) \\ &= \min\{t - \pi_{m-i}, t - \pi_{m-k}\}. \end{aligned}$$

Example 3.3. Let $X = \{X_t : 0 \leqslant t \leqslant T\}$ be the unique strong solution of equation (3.1) and Π a collection of delay points with just two points. Then, we have

$$\mathbf{X}_{\Pi_t} = (X_t^{\pi}, X_t), \qquad 0 \leqslant t \leqslant T,$$

where

$$X_t = X_0 + \int_0^t b(s, X_s) ds + \int_0^t \sigma(s, X_s) dW_s, \qquad 0 \leqslant t \leqslant T, \qquad \text{a.s.},$$

and

$$X_t^{\pi} = X_{\pi}^{\pi} + \int_{\pi}^t b(s-\pi, X_s^{\pi}) ds + \int_{\pi}^t \sigma(s-\pi, X_s^{\pi}) dW_s^{\pi}, \qquad \pi \leqslant t \leqslant T, \qquad \text{a.s.}$$

Example 3.4. Let $X = \{X_t : 0 \leqslant t \leqslant T\}$ be the unique strong solution of equation (3.1) and Π given by $\Pi = \{0 = \pi_0 < \pi_1 < \cdots < \pi_m = \pi\}$. Then the process $\mathbf{X}_\Pi = \{\mathbf{X}_{\Pi_t} : \pi \leqslant t \leqslant T\}$ is given by

$$\mathbf{X}_{\Pi_t} = (X_t^{\pi_m}, X_t^{\pi_{m-1}}, \cdots, X_t^{\pi_1}, X_t), \qquad \pi \leqslant t \leqslant T,$$

where

$$\begin{aligned} X_t^{\pi_{m-(i-1)}} = X_{\pi_{m-(i-1)}}^{\pi_{m-(i-1)}} &+ \int_{\pi_{m-(i-1)}}^t b(s - \pi_{m-(i-1)}, X_s^{\pi_{m-(i-1)}}) ds \\ &+ \int_{\pi_{m-(i-1)}}^t \sigma(s - \pi_{m-(i-1)}, X_s^{\pi_{m-(i-1)}}) dW_s^{\pi_{m-(i-1)}}, \end{aligned}$$

for all $\pi_{m-(i-1)} \leqslant t \leqslant T$ and $i = 1, \cdots, m+1$, a.s.

Let us introduce the $(m+1)$-dimensional vector

$$\mathbf{X}_{\Pi_0} = (X_0, \cdots, X_0),$$

and the functions

$$\mathbf{b} : [0,T] \times \mathbb{R}^{m+1} \to \mathbb{R}^{m+1} \quad \text{and} \quad \boldsymbol{\sigma} : [0,T] \times \mathbb{R}^{m+1} \to \mathbb{R}^{(m+1)\times(m+1)},$$

given by

$$\mathbf{b} = (b_i)_{i=1}^{m+1} \quad \text{with} \quad b_i(s, \mathbf{x}) = b(s - \pi_{m-(i-1)}, x^{(m-i)}) 1_{[\pi_{m-(i-1)}, \infty)}(s),$$

and

$$\boldsymbol{\sigma} = (\sigma_{ij})_{i,j=1}^{m+1} \quad \text{with} \quad \sigma_{ij}(s, \mathbf{x}) = \begin{cases} \sigma(s - \pi_{m-(i-1)}, x^{(m-i)}) 1_{[\pi_{m-(i-1)}, \infty)}(s) & \text{if } i = j, \\ 0 & \text{if } i \neq j, \end{cases}$$

respectively. By using these quantities, we can write

$$\mathbf{X}_{\Pi_t}{}^{(i)} = \mathbf{X}_{\Pi_0}{}^{(i)} + \int_0^t b_i(s, \mathbf{X}_{\Pi_s})ds + \sum_{j=1}^{m+1} \int_0^t \sigma_{ij}(s, \mathbf{X}_{\Pi_s})d\boldsymbol{W}_{\Pi_s}{}^{(j)} \quad 0 \leqslant t \leqslant T, a.s.,$$

for $i = 1, \cdots, m+1$; or briefly

$$\mathbf{X}_{\Pi_t} = \mathbf{X}_{\Pi_0} + \int_0^t \mathbf{b}(s, \mathbf{X}_{\Pi_s})ds + \int_0^t \boldsymbol{\sigma}(s, \mathbf{X}_{\Pi_s})d\boldsymbol{W}_{\Pi_s}, \tag{3.8}$$

where $\boldsymbol{W}_\Pi$ is the delay vector corresponding to the Brownian motion and the collection of delay points Π. Notice that the notation $d\boldsymbol{W}_{\Pi_s}$ in the last integral of equation (3.8) is only formal as each component has to be considered w.r.t. a different filtration. So one cannot choose one filtration and consider the equation w.r.t. this filtration as one usually does in the context of SDEs.

3.3 Itô Type Formula for Delay Vectors

Now, we study processes of the form (3.4) when X is a solution of equation (3.1).

3.3.1 The Brownian Motion Case

The next theorem would be covered by Theorem 2.1 in Hu, Mohammed, and Yan (2004) but giving a different proof. First, we will focus on the case when X is simply a standard Brownian motion W which allows us to illustrate the main ideas behind the general case as well as some of the main difficulties of the proof.

Theorem 3.3. *(Itô Type Formula for Delay Vectors: Brownian Motion Case) Let W be a standard Brownian motion and $\Pi = \{0 = \pi_0 < \cdots < \pi_m = \pi\}$ a collection of delay points. Consider the process $\boldsymbol{W}_\Pi$, and a continuous function $f : \mathbb{R}^{m+1} \to \mathbb{R}$. If f*

belongs to $C_p^3(\mathbb{R}^{m+1})$*, for some* $p \geqslant 1$ *then the process* $f(\boldsymbol{W}_{\Pi})$ *given by* $\{f(\boldsymbol{W}_{\Pi_t}) : t \geqslant \pi\}$ *satisfies*

$$f(\boldsymbol{W}_{\Pi_t}) = f(\boldsymbol{W}_{\Pi_\pi}) + \sum_{j=1}^{m+1} \int_{\pi-\pi_{(m+1-j)}}^{t-\pi_{(m+1-j)}} \partial_j f(\boldsymbol{W}_{\Pi_{s+\pi_{(m+1-j)}}}) \delta W_s$$
$$+ \frac{1}{2} \sum_{j=1}^{m+1} \sum_{k=1}^{m+1} \int_{\pi}^{t} \partial_{jk}^2 f(\boldsymbol{W}_{\Pi_s}) ds, \qquad \text{a.s.},$$

where

$$\int_{\pi-\pi_{(m+1-j)}}^{t-\pi_{(m+1-j)}} \partial_j f(\boldsymbol{W}_{\Pi_{s+\pi_{(m+1-j)}}}) \delta W_s,$$

is a stochastic integral, with respect to the Brownian motion, in the Skorohod sense (see Section 2.4 for details about the Skorohod integral).

Proof. Let $t \geqslant \pi$ be a fixed time and

$$P^N = \{\pi = s_0^N < s_1^N < \cdots < s_N^N = t\}, \qquad N \geqslant 1,$$

be a sequence of partitions of the interval $[\pi, t]$ such that $|P^N|$ tends to zero as N goes to infinity. Besides, let us assume that

$$|P^N| = \max_i \{s_{i+1}^N - s_i^N\} \leqslant \pi, \qquad \text{for all} \quad N \geqslant 1. \tag{3.9}$$

We can make this assumption without loss of generality since the mesh tends to zero as N goes to infinity. Hereafter, we will write s_i instead of s_i^N for notational convenience. For any $N \geqslant 1$, we can write

$$f(\boldsymbol{W}_{\Pi_t}) - f(\boldsymbol{W}_{\Pi_\pi}) = \sum_{i=1}^{N} [f(\boldsymbol{W}_{\Pi_{s_i}}) - f(\boldsymbol{W}_{\Pi_{s_{i-1}}})],$$

which gives, when using Taylor's formula up to the second order,

$$
\begin{aligned}
f(\boldsymbol{W}_{\Pi_t}) - f(\boldsymbol{W}_{\Pi_\pi}) &= \sum_{i=1}^{N} Df(\boldsymbol{W}_{\Pi_{s_{i-1}}})(\boldsymbol{W}_{\Pi_{s_i}} - \boldsymbol{W}_{\Pi_{s_{i-1}}}) \\
&\quad + \frac{1}{2}\sum_{i=1}^{N} (\boldsymbol{W}_{\Pi_{s_i}} - \boldsymbol{W}_{\Pi_{s_{i-1}}})^{T} D^{(2)} f(\xi_{\Pi_i})(\boldsymbol{W}_{\Pi_{s_i}} - \boldsymbol{W}_{\Pi_{s_{i-1}}}),
\end{aligned}
$$

where ξ_{Π_i} is a random intermediate point between $\boldsymbol{W}_{\Pi_{s_i}}$ and $\boldsymbol{W}_{\Pi_{s_{i-1}}}$, i.e.

$$
\xi_{\Pi_i} = \theta \boldsymbol{W}_{\Pi_{s_{i-1}}} + (1-\theta)\boldsymbol{W}_{\Pi_{s_i}},
$$

for some random θ in $(0,1)$. So, we can write

$$
f(\boldsymbol{W}_{\Pi_t}) - f(\boldsymbol{W}_{\Pi_\pi}) = S(P^N, \Pi) + R(P^N, \Pi), \tag{3.10}
$$

where

$$
S(P^N, \Pi) = \sum_{j=1}^{m+1}\sum_{i=1}^{N} \partial_j f(\boldsymbol{W}_{\Pi_{s_{i-1}}})(W_{s_i - \pi_{m-(j-1)}} - W_{s_{i-1}-\pi_{m-(j-1)}}),
$$

and

$$
\begin{aligned}
R(P^N, \Pi) = \frac{1}{2}\sum_{j=1}^{m+1}\sum_{k=1}^{m+1}\sum_{i=1}^{N} &(W_{s_i - \pi_{m-(j-1)}} - W_{s_{i-1}-\pi_{m-(j-1)}}) \\
&\partial^2_{jk} f(\xi_{\Pi_i})(W_{s_i - \pi_{m-(k-1)}} - W_{s_{i-1}-\pi_{m-(k-1)}}).
\end{aligned}
$$

Next, we investigate the convergence of these sums as N tends to infinity. In order to keep clarity we are going to decompose the proof in several steps.

Step 1. We shall prove that $S(P^N, \Pi)$ converges in $L^2(\Omega)$-norm as N tends to infinity. To do this, let us first introduce some ancillary quantities. Let

$$
S_j(P^N, \Pi) = \sum_{i=1}^{N} F_i^j W(1_{A_i^j}). \qquad j = 1, \cdots, m+1,
$$

where

$$F_i^j = \partial_j f(\boldsymbol{W}_{\boldsymbol{\Pi}_{s_{i-1}}}),$$

and

$$1_{A_i^j}(s) = 1_{(s_{i-1}-\pi_{m-(j-1)}, s_i-\pi_{m-(j-1)}]}(s).$$

Using these quantities, we can write

$$S(P^N, \Pi) = \sum_{j=1}^{m+1} S_j(P^N, \Pi). \tag{3.11}$$

In addition, for each $N \geqslant 1$, let

$$u_j^N = \sum_{i=1}^{N} F_i^j 1_{A_i^j}, \qquad j = 1, \cdots, m+1. \tag{3.12}$$

Now, as $f \in C_p^3(\mathbb{R}^{m+1})$, we have $F_i^j \in \mathbb{D}^{1,2}$ for all $i = 1, \cdots, N, j = 1, \cdots, m+1$. Hence, Proposition 2.7 and the linearity of the Skorohod integral imply

$$\begin{aligned}\delta(u_j^N) &= \sum_{i=1}^{N} F_i^j W(1_{A_i^j}) - \sum_{i=1}^{N} \langle DF_i^j, 1_{A_i^j}\rangle_H \\ &= S_j(P^N, \Pi) - \sum_{i=1}^{N} \int_0^t D_s F_i^j 1_{A_i^j}(s) ds, \qquad a.s.\end{aligned}$$

That is,

$$S_j(P^N, \Pi) = \delta(u_j^N) + \int_0^t \sum_{i=1}^{N} D_s F_i^j 1_{A_i^j}(s) ds, \qquad a.s. \tag{3.13}$$

So, the convergence of $S_j(P^N, \Pi)$ (and consequently the convergence of $S(P^N, \Pi)$) can be determined by studying the convergence of the terms on the right-hand side of this equality.

Step 2. In order to study the first term on the right hand side of (3.13) , let us define the following ancillary quantities

$$u_j^\infty(s) = F^j \nu^j(s) \qquad s \geqslant 0,$$

where

$$F^j(s) = \partial_j f(\boldsymbol{W}_{\Pi_{s+\pi_{m-(j-1)}}}), \qquad s \geqslant 0,$$

and

$$\nu^j(s) = 1_{(\pi-\pi_{m-(j-1)}, t-\pi_{m-(j-1)}]}(s), \qquad s \geqslant 0.$$

Next, we are going to verify that

$$u_j^N \to u_j^\infty \qquad \text{as} \qquad N \to \infty, \qquad \text{in } \mathbb{D}^{1,2}(H), \tag{3.14}$$

where $H = L^2([0,t])$ with t being the time fixed at the beginning of the proof. So, we fix $1 \leqslant j \leqslant m+1$ and estimate

$$\mathbb{E}\left[\int_0^t |u_j^\infty(s) - u_j^N(s)|^2 ds\right] + \mathbb{E}\left[\int_0^t \int_0^t |D_r u_j^\infty(s) - D_r u_j^N(s)|^2 dr ds\right]. \tag{3.15}$$

Using the definitions of u_j^∞ and u_j^N, the first summand reads

$$\begin{aligned}
\mathbb{E}\left[\int_0^t |u_j^\infty(s) - u_j^N(s)|^2 ds\right] &= \sum_{i=1}^N \int_{s_{i-1}}^{s_i} \mathbb{E}\left[|\partial_j f(\boldsymbol{W}_{\Pi_s}) - \partial_j f(\boldsymbol{W}_{\Pi_{s_{i-1}}})|^2\right] ds \\
&\leqslant C \sum_{i=1}^N \sum_{l=1}^{m+1} \left\{ \left(\int_{s_{i-1}}^{s_i} \mathbb{E}[|\partial_{jl}^2 f(\theta \boldsymbol{W}_{\Pi_s} + (1-\theta)\boldsymbol{W}_{\Pi_{s_{i-1}}})|^4] ds \right)^{1/2} \right. \\
&\qquad \left. \left(\int_{s_{i-1}}^{s_i} \mathbb{E}[|W_{s-\pi_{m-l+1}} - W_{s_{i-1}-\pi_{m-l+1}}|^4] ds \right)^{1/2} \right\},
\end{aligned}$$

by mean value theorem followed by Cauchy-Schwarz inequality, where $\theta \in [0,1]$ is a random number.

Now, notice that as $f \in C_p^3(\mathbb{R}^{m+1})$ there exists a power $p \geqslant 1$ such that

$$\left|\partial_{jl}^2 f(\theta \boldsymbol{W}_{\Pi_s} + (1-\theta)\boldsymbol{W}_{\Pi_{s_{i-1}}})\right| \leqslant C\left(1 + |\theta \boldsymbol{W}_{\Pi_s} + (1-\theta)\boldsymbol{W}_{\Pi_{s_{i-1}}}|^p\right),$$

where

$$\mathbb{E}\left[\left|\theta \boldsymbol{W}_{\Pi_s} + (1-\theta)\boldsymbol{W}_{\Pi_{s_{i-1}}}\right|^{2p}\right] \leqslant C t^p,$$

by Gaussian integration using $\theta \in [0,1]$ and $s_{i-1} \leqslant s \leqslant s_i \leqslant t$. Therefore

$$\mathbb{E}\left[|\partial^2_{jl} f(\theta \boldsymbol{W}_{\Pi_s} + (1-\theta)\boldsymbol{W}_{\Pi_{s_{i-1}}})|^4\right] \leqslant C_f(1+t^{2p}), \tag{3.16}$$

where the constant C_f only depends on m and p, that is, on the function f.

Thus,

$$\left(\int_{s_{i-1}}^{s_i} \mathbb{E}\left[|\partial^2_{jl} f(\theta \boldsymbol{W}_{\Pi_s} + (1-\theta)\boldsymbol{W}_{\Pi_{s_{i-1}}})|^4\right] ds\right)^{1/2} \leqslant C(s_i - s_{i-1})^{1/2},$$

for all $i = 1, \cdots, N$, and $l = 1, \cdots, m+1$. Furthermore, again by Gaussian integration, we obtain

$$\begin{aligned}\left(\int_{s_{i-1}}^{s_i} \mathbb{E}[|W_{s-\pi_{m-l+1}} - W_{s_{i-1}-\pi_{m-l+1}}|^4]ds\right)^{1/2} &\leqslant C\left(\int_{s_{i-1}}^{s_i}(s-s_{i-1})^2 ds\right)^{1/2}\\ &= \frac{C}{\sqrt{3}}(s_i - s_{i-1})^{3/2},\end{aligned}$$

for all $i = 1, \cdots, N$, and $l = 1, \cdots, m+1$, and hence the first summand of equation (3.15) is bounded by

$$C(m+1)\sum_{i=1}^{m+1}(s_i - s_{i-1})^{1/2}(s_i - s_{i-1})^{3/2},$$

which converges to zero when $N \to \infty$ as

$$\sum_{i=1}^{N}(s_i - s_{i-1}) = t - \pi.$$

Working out $D_r u_j^\infty(s)$ and $D_r u_j^N(s)$ explicitly in the second sum of (3.15) yields

$$\begin{aligned}
&\sum_{i=1}^{N}\mathbb{E}\left[\int_{s_{i-1}}^{s_i}\int_0^t\left(\sum_{k=1}^{m+1}\left\{\partial_{kj}^2 f(\boldsymbol{W}_{\Pi_s})1_{[0,s-\pi_{m-k+1}]}(r)-\partial_{kj}^2 f(\boldsymbol{W}_{\Pi_{s_{i-1}}})1_{[0,s_{i-1}-\pi_{m-k+1}]}(r)\right\}\right)^2 drds\right]\\
&\quad\leqslant C\sum_{k=1}^{m+1}\sum_{i=1}^{N}\int_{s_{i-1}}^{s_i}\mathbb{E}\left[\left|\partial_{kj}^2 f(\boldsymbol{W}_{\Pi_s})\right|^2\right](s-s_{i-1})ds\\
&\qquad+C\sum_{k=1}^{m+1}\sum_{i=1}^{N}\int_{s_{i-1}}^{s_i}\mathbb{E}\left[|\partial_{kj}^2 f(\boldsymbol{W}_{\Pi_s})-\partial_{kj}^2 f(\boldsymbol{W}_{\Pi_{s_{i-1}}})|^2\right](s_{i-1}-\pi_{m-k+1})ds.
\end{aligned}$$

By the same argument used to derivate (3.16),

$$\mathbb{E}\left[\left|\partial_{kj}^2 f(\boldsymbol{W}_{\Pi_s})\right|^2\right]\leqslant C(1+t^p),$$

and hence the first sum on the above right hand side is bounded by

$$C(m+1)\sum_{i=1}^{N}(s_i-s_{i-1})^2\to 0,\qquad N\to\infty. \tag{3.17}$$

For showing the convergence of the second sum, it suffices to estimate

$$\sum_{i=1}^{N}\int_{s_{i-1}}^{s_i}\mathbb{E}\left[\left|\partial_{kj}^2 f(\boldsymbol{W}_{\Pi_s})-\partial_{kj}^2 f(\boldsymbol{W}_{\Pi_{s_{i-1}}})\right|^2\right]ds, \tag{3.18}$$

for fixed $j=1,\cdots,m+1$, since $|s_{i-1}-\pi_{k-m+1}|\leqslant t$, for all $i=1,\cdots,N$ and all $k=1,\cdots,m+1$. However, the convergence of this sum can be shown in almost exactly the same way the convergence to zero when $N\to\infty$ was shown for the first summand of equation (3.15). The only difference is that $\partial_j f$ has to be replaced by $\partial_{kj}^2 f$, but the mean value theorem can still be used because $f\in C_p^3(\mathbb{R}^{m+1})$.

All in all, both summands of (3.15) converge to zero when $N\to\infty$ which finally proves (3.14).

Now, by using the convergence in (3.14) together with Proposition 2.6 we can conclude that the sequence $(\delta(u_j^N))_{N\geqslant 1}$ converges in $L^2(\Omega)$ to

$$\int_0^t \partial_j f(\boldsymbol{W}_{\Pi_{s+\pi_{m-(j-1)}}}) 1_{(\pi-\pi_{m-(j-1)},t-\pi_{m-(j-1)}]}(s)\delta W_s,$$

as N tends to infinity, which in turn implies that

$$\sum_{j=1}^{m+1} \delta(u_j^N) \to \sum_{j=1}^{m+1} \int_0^t \partial_j f(\boldsymbol{W}_{\Pi_{s+\pi_{m-j+1}}}) 1_{(\pi-\pi_{m+1-j},t-\pi_{m+1-j}]}(s)\delta W_s, \tag{3.19}$$

in $L^2(\Omega)$ as N tends to infinity.

Step 3. For the last term on the right hand side of (3.13) notice that

$$\begin{aligned}\sum_{i=1}^{N} &\langle DF_i^j, 1_{A_i^j}\rangle_H \\ &= \sum_{i=1}^{N} \int_0^t \sum_{k=1}^{m+1} \partial_{kj}^2 f(\boldsymbol{W}_{\Pi_{s_{i-1}}}) 1_{[0,s_{i-1}-\pi_{m+1-k}]}(s) 1_{(s_{i-1}-\pi_{m-(j-1)},s_i-\pi_{m-(j-1)}]}(s)ds,\end{aligned}$$

that is,

$$\sum_{i=1}^{N} \langle DF_i^j, 1_{A_i^j}\rangle_H = \sum_{i=1}^{N} \sum_{k=1}^{m+1} \int_0^t \partial_{kj}^2 f(\boldsymbol{W}_{\Pi_{s_{i-1}}}) I_{kj}^{i-1}(s)ds,$$

where I_{kj}^{i-1} is the indicator function of the intersection of the intervals $[0, s_{i-1} - \pi_{m+1-k}]$ and $(s_{i-1} - \pi_{m-(j-1)}, s_i - \pi_{m-(j-1)}]$.

Of course

$$I_{kj}^{i-1} = \begin{cases} 1_\emptyset & \text{if } m-(j-1) \leqslant m+1-k, \\ 1_{(s_{i-1}-\pi_{m-(j-1)},s_i-\pi_{m-(j-1)}]} & \text{if } m-(j-1) > m+1-k, \end{cases}$$

so that, due to (3.9), we obtain

$$\sum_{i=1}^{N} \langle DF_i^j, 1_{A_i^j}\rangle_H = \sum_{i=1}^{N} \sum_{k=j+1}^{m+1} \int_0^t \partial_{kj}^2 f(\boldsymbol{W}_{\Pi_{s_{i-1}}}) 1_{(s_{i-1}-\pi_{m-(j-1)},s_i-\pi_{m-(j-1)}]}(s)ds \cdot 1_{\{j\leqslant m\}}.$$

Therefore, using the convergence of equation (3.18), we have that

$$\sum_{i=1}^{N}\langle DF_i^j, 1_{A_i^j}\rangle_H \to \sum_{k=j+1}^{m+1}\int_{\pi-\pi_{m+1-j}}^{t-\pi_{m+1-j}} \partial_{kj}^2 f(\mathbf{W}_{\Pi_{s+\pi_{m+1-j}}})ds \cdot 1_{\{j\leqslant m\}}, \tag{3.20}$$

in $L^2(\Omega)$, as N tends to infinity.

Step 4. Summing up, equations (3.19) and (3.49) imply that

$$\begin{aligned} S(P^N,\Pi) \to &\sum_{j=1}^{m+1}\int_0^\infty \partial_j f(\mathbf{W}_{\Pi_{s+\pi_{m-j+1}}})1_{[\pi-\pi_{m+1-j},t-\pi_{m+1-j}]}(s)\delta W_s \\ &+\sum_{k>j}\int_{\pi-\pi_{m+1-j}}^{t-\pi_{m+1-j}} \partial_{kj}^2 f(\mathbf{W}_{\Pi_{s+\pi_{m+1-j}}})ds, \end{aligned} \tag{3.21}$$

in $L^2(\Omega)$, as N tends to infinity.

Step 5. Next, we study the convergence of $R(P^N,\Pi)$ which is the second term on the right-hand side of equation (3.10). First , we write

$$R(P^N,\Pi) = \frac{1}{2}\sum_{j=1}^{m+1}\sum_{k=1}^{m+1}\tilde{R}_{jk}(P^N,\Pi), \tag{3.22}$$

where

$$\tilde{R}_{jk}(P^N,\Pi) = \sum_{i=1}^{N}\partial_{jk}^2 f(\xi_{\Pi_i})(W_{s_i-\pi_{m-(k-1)}} - W_{s_{i-1}-\pi_{m-(k-1)}})(W_{s_i-\pi_{m-(j-1)}} - W_{s_{i-1}-\pi_{m-(j-1)}}),$$

and note that, under the limit, $\tilde{R}_{jk}(P^N,\Pi)$ can be replaced by

$$R_{jk}(P^N,\Pi) = \sum_{i=1}^{N}\partial_{jk}^2 f(\mathbf{W}_{\Pi_{s_{i-1}}})(W_{s_i-\pi_{m-(k-1)}} - W_{s_{i-1}-\pi_{m-(k-1)}})(W_{s_i-\pi_{m-(j-1)}} - W_{s_{i-1}-\pi_{m-(j-1)}})$$

because

$$\mathbb{E}\left[\left|\tilde{R}_{jk}(P^N,\Pi) - R_{jk}(P^N,\Pi)\right|^2\right] \to 0, \qquad N\to\infty, \tag{3.23}$$

for all $1 \leqslant j, k \leqslant m+1$. Indeed,

$$\mathbb{E}\left[\sum_{i=1}^{N}\Big|\partial^2_{jk}f(\theta \boldsymbol{W}_{\Pi_{s_{i-1}}}+(1-\theta)\boldsymbol{W}_{\Pi_{s_i}})-\partial^2_{jk}f(\boldsymbol{W}_{\Pi_{s_{i-1}}})\right.$$

$$\left.(W_{s_i-\pi_{m-(k-1)}}-W_{s_{i-1}-\pi_{m-(k-1)}})(W_{s_i-\pi_{m-(j-1)}}-W_{s_{i-1}-\pi_{m-(j-1)}})\Big|^2\right]$$

$$=\mathbb{E}\left[\left|\sum_{i=1}^{N}\sum_{l=1}^{m+1}\partial^3_{jkl}f(\tilde{\xi}_{\Pi_i})(1-\tilde{\theta})(W_{s_i-\pi_{m-(l-1)}}-W_{s_{i-1}-\pi_{m-(l-1)}})\right.\right.$$

$$\left.\left.(W_{s_i-\pi_{m-(k-1)}}-W_{s_{i-1}-\pi_{m-(k-1)}})(W_{s_i-\pi_{m-(j-1)}}-W_{s_{i-1}-\pi_{m-(j-1)}})\right|^2\right],$$

where

$$\tilde{\xi}_{\Pi_i}=\tilde{\theta}\boldsymbol{W}_{\Pi_{s_{i-1}}}+(1-\tilde{\theta})[\theta\boldsymbol{W}_{\Pi_{s_{i-1}}}+(1-\theta)\boldsymbol{W}_{\Pi_{s_i}}],$$

for some random $\tilde{\theta}\in[0,1]$, by mean value theorem, and hence (3.23) holds true if, for each $1 \leqslant j, k, l \leqslant m+1$,

$$\sum_{i=1}^{N}\sum_{i'=1}^{N}\mathbb{E}\,[|\ \partial^3_{jkl}f(\tilde{\xi}_{\Pi_i})\partial^3_{jkl}f(\tilde{\xi}_{\Pi_{i'}})$$

$$(W_{s_i-\pi_{m-(l-1)}}-W_{s_{i-1}-\pi_{m-(l-1)}}(W_{s_{i'}-\pi_{m-(l-1)}}-W_{s_{i'-1}-\pi_{m-(l-1)}})$$

$$(W_{s_i-\pi_{m-(k-1)}}-W_{s_{i-1}-\pi_{m-(k-1)}}(W_{s_{i'}-\pi_{m-(k-1)}}-W_{s_{i'-1}-\pi_{m-(k-1)}})$$

$$(W_{s_i-\pi_{m-(j-1)}}-W_{s_{i-1}-\pi_{m-(j-1)}}(W_{s_{i'}-\pi_{m-(j-1)}}-W_{s_{i'-1}-\pi_{m-(j-1)}})|\,],$$

converges to zero when $N\to\infty$. But by Gaussian integration after successively applying Cauchy-Schwarz' inequality, using the assumption $f\in C^3_p(\mathbb{R}^{m+1})$ as in the proof of (3.16), the summands in the above sum can be estimated by

$$C(1+t^p)(s_i-s_{i-1})^{3/2}(s_{i'}-s_{i'-1})^{3/2},$$

for some $p \geqslant 1$, uniformly in i and i', so that this sum converges to zero when $N\to\infty$ as

$$\sum_{i=1}\sum_{i'=1}^{N}(s_i-s_{i-1})(s_{i'}-s_{i'-1})=(t-\pi)^2.$$

In what follows, we therefore work with $R_{jk}(P^N, \Pi)$ instead of $\tilde{R}_{jk}(P^N, \Pi)$.

Step 6. For each $N \geqslant 1$, let

$$v_{jk}^N = \sum_{i=1}^N G_i^{jk,k} 1_{A_i^j}, \qquad j,k = 1, \cdots, m+1, \tag{3.24}$$

where

$$G_i^{jk,l} = \frac{1}{2} \partial_{jk}^2 f(\boldsymbol{W}_{\Pi_{s_{i-1}}})(W_{s_i - \pi_{m-(l-1)}} - W_{s_{i-1} - \pi_{m-(l-1)}}),$$

and, as before,

$$1_{A_i^j} = 1_{(s_{i-1} - \pi_{m-(j-1)}, s_i - \pi_{m-(j-1)}]}.$$

In this notation, since $f \in C_p^3(\mathbb{R}^{m+1})$, we have $G_i^{jk,k} \in \mathbb{D}^{1,2}$ for all $i = 1, \cdots, N$, $1 \leqslant j, k \leqslant m+1$, and hence

$$\delta(v_{jk}^N) = \sum_{i=1}^N G_i^{jk,k} W(1_{A_i^j}) - \sum_{i=1}^N \langle DG_i^{jk,k}, 1_{A_i^j} \rangle_H, \qquad \text{a.s.}, \tag{3.25}$$

by Proposition 2.7 and linearity of the Skorohod integral.

Now, as

$$\frac{1}{2} R_{jk}(P^N, \Pi) = \sum_{i=1}^N G_i^{jk,k} W(1_{A_i^j}),$$

equation (3.25) implies

$$\frac{1}{2} R_{jk}(P^N, \Pi) = \delta(v_{jk}^N) + \sum_{i=1}^N \langle DG_i^{jk,k}, 1_{A_i^j} \rangle_H, \qquad \text{a.s.}, \qquad j,k = 1, \cdots, m+1,$$

which in turn gives

$$R(P^N, \Pi) = \sum_{j=1}^{m+1} \sum_{k=1}^{m+1} \delta(v_{jk}^N) + \sum_{j=1}^{m+1} \sum_{k=1}^{m+1} \sum_{i=1}^N \langle DG_i^{jk,k}, 1_{A_i^j} \rangle_H, \qquad \text{a.s.} \tag{3.26}$$

using $R(P^N, \Pi)$ instead of $\tilde{R}(P^N, \Pi)$ as justified in Step 5.

So, it suffices to study the convergence of the two sums on the above right-hand side.

Step 7. We claim that the first sum converges to zero in $L^2(\Omega)$ when $N \to \infty$. To see this we apply Proposition 2.6 and show that

$$v_{jk}^N \to 0, \qquad N \to \infty, \qquad \text{in } \mathbb{D}^{1,2}(H),$$

for arbitrary but fixed $1 \leqslant j,k \leqslant m+1$, where $H = L^2([0,t])$ as in Step 2. The task is to estimate

$$\mathbb{E}\left[\int_0^t \left|\sum_{i=1}^N G_i^{jk,k} 1_{A_i^j}(s)\right|^2 ds\right] + \mathbb{E}\left[\int_0^t \int_0^t \left|\sum_{i=1}^N D_r G_i^{jk,k} 1_{A_i^j}(s)\right|^2 drds\right], \tag{3.27}$$

where the first summand simplifies to

$$\sum_{i=1}^N \mathbb{E}\left[\left|\frac{1}{2}\partial_{jk}^2 f(\boldsymbol{W}_{\Pi_{s_{i-1}}})\right|^2 |W_{s_i-\pi_{m-k+1}} - W_{s_{i-1}-\pi_{m-k+1}}|^2\right](s_i - s_{i-1})$$
$$\leqslant C\sum_{i=1}^N (1+t^p)(s_i - s_{i-1})^2,$$

using Cauchy-Schwarz inequality, followed by Gaussian integration — see the proof of equation (3.16) for more details leading to the term $(1+t^p)$. Thus, the first summand of (3.27) converges to zero when $N \to \infty$ since $\sum_{i=1}^N (s_i - S_{i-1}) = t - \pi$.

The second summand equals

$$\mathbb{E}\Bigg[\int_0^t \int_0^t \Bigg|\sum_{i=1}^N \sum_{l=1}^{m+1} \frac{1}{2}\partial_{jkl}^3 f(\boldsymbol{W}_{\Pi_{s_{i-1}}}) 1_{[0,s_{i-1}-\pi_{m-l+1}]}(r)(W_{s_i-\pi_{m-k+1}} - W_{s_{i-1}-\pi_{m-k+1}}) 1_{A_i^j}(s)$$
$$+ \sum_{i=1}^N \frac{1}{2}\partial_{jk}^2 f(\boldsymbol{W}_{\Pi_{s_{i-1}}}) 1_{(s_{i-1}-\pi_{m-k+1}, s_i-\pi_{m-k+1}]}(r) 1_{A_i^j}(s)\Bigg|^2 drds\Bigg],$$

which can be bounded by

$$C\sum_{i=1}^{N}\left[(1+t^p)(m+1)t(s_{i-1}-s_i)^2+(1+t^p)(s_{i-1}-s_i)^2\right],$$

applying the same ideas used to bound the first summand of (3.27), but also using the fact that

$$\int_0^t 1_{[0,s_{i-1}-\pi_{m-l+1}]}(r)dr \leqslant t, \qquad \text{for all } 1\leqslant i\leqslant N, 1\leqslant l\leqslant m+1.$$

Note that, without restricting the generality, we assume the same polynomial growth for both $\partial^2_{jk}f$ and $\partial^3_{jkl}f$.

All in all, the second sum of (3.27) converges to zero when $N\to\infty$, too, so that indeed v^N_{jk} converges to zero as $N\to\infty$, in $\mathbb{D}^{1,2}(H)$, for any $1\leqslant jk\leqslant m+1$.

Step 8. To study the triple sum in equation (3.26) we fix $1\leqslant j,k\leqslant m+1$ and split

$$\sum_{i=1}^{N}\langle DG_i^{jk,k}, 1_{A_i^j}\rangle_H,$$

into two sums:

$$\sum_{i=1}^{N}\sum_{l=1}^{m+1}\int_0^t \frac{1}{2}\partial^3_{jkl}f(\boldsymbol{W}_{\Pi_{s_{i-1}}})1_{[0,s_{i-1}-\pi_{m-l+1}]}(r)(W_{s_i-\pi_{m-k+1}}-W_{s_{i-1}-\pi_{m-k+1}})1_{A_i^j}(r)dr,$$

and

$$\sum_{i=1}^{N}\int_0^t \frac{1}{2}\partial^2_{jk}f(\boldsymbol{W}_{\Pi_{s_{i-1}}})1_{A_i^k\cap A_i^j}(r)dr. \tag{3.28}$$

First, we claim the first sum converges to zero, in $L^2(\mathbb{P})$, when N goes to infinity. To see this, observe that the $L^2(\mathbb{P})$-norm square of the first sum can be bounded by

$$
\begin{aligned}
&(m+1)t\sum_{l=1}^{m+1}\int_0^t \mathbb{E}\left[\left|\sum_{i=1}^{N}\frac{1}{2}\left|\partial^3_{jkl}f(W_{\Pi_{s_{i-1}}})\right|\left|W_{s_i-\pi_{m-k+1}}-W_{s_{i-1}-\pi_{m-k+1}}\right|1_{A_i^j}(r)\right|^2\right]dr\\
&=(m+1)t\sum_{l=1}^{m+1}\sum_{i=1}^{N}\mathbb{E}\left[\left|\frac{1}{2}\partial^3_{jkl}f(W_{\Pi_{s_{i-1}}})\right|^2\left|W_{s_i-\pi_{m-k+1}}-W_{s_{i-1}-\pi_{m-k+1}}\right|^2\right](s_i-s_{i-1})\\
&\leqslant (m+1)t\sum_{l=1}^{m+1}\sum_{i=1}^{N}C(1+t^p)(s_i-s_{i-1})(s_i-s_{i-1}),
\end{aligned}
$$

using Cauchy-Schwarz inequality followed by Gaussian integration just as for the estimate below equation (3.27). But, the last sum of course converges to zero when $N\to\infty$.

Second, looking at equation (3.28), one can notice that when N is large enough, for any $j\neq k$ the two sets A_i^k and A_i^j are disjoint for all $i=1,\cdots,N$. Therefore, the sum associated with (3.28) trivially vanishes whenever $j\neq k$. Now, considering the case $j=k$, we claim that

$$
\sum_{i=1}^{N}\int_0^t\frac{1}{2}\partial^2_{jj}f(W_{\Pi_{s_{i-1}}})1_{A_i^j}(r)dr\to\frac{1}{2}\int_{\pi-\pi_{m-j+1}}^{t-\pi_{m-j+1}}\partial^2_{jj}f(W_{\Pi_{r+\pi_{m-j+1}}})dr,
$$

in $L^2(\mathbb{P})$, as $N\to\infty$.

Indeed, because

$$
\begin{aligned}
&\mathbb{E}\left[\left|\int_0^t\left\{\frac{1}{2}\partial^2_{jj}f(W_{\Pi_{r+\pi_{m-j+1}}})1_{(\pi-\pi_{m-j+1},t-\pi_{m-j+1}]}(r)-\sum_{i=1}^{N}\frac{1}{2}\partial^2_{jj}f(W_{\Pi_{s_{i-1}}})1_{A_i^j}(r)\right\}dr\right|^2\right]\\
&\leqslant t\mathbb{E}\left[\int_0^t\left|\frac{1}{2}\partial^2_{jj}f(W_{\Pi_{r+\pi_{m-j+1}}})1_{(\pi-\pi_{m-j+1},t-\pi_{m-j+1}]}(r)-\sum_{i=1}^{N}\frac{1}{2}\partial^2_{jj}f(W_{\Pi_{s_{i-1}}})1_{A_i^j}(r)\right|^2dr\right],
\end{aligned}
$$

the claim follows by the same arguments used to show the convergence of the first summand of (3.15).

All in all

$$\sum_{i=1}^{N}\sum_{j=1}^{m+1}\sum_{k=1}^{m+1}\langle DG_i^{jk,k}, 1_{A_i^j}\rangle_H \to \frac{1}{2}\sum_{j=1}^{m+1}\int_{\pi-\pi_{m+1-j}}^{t-\pi_{m+1-j}} \partial_{jj}^2 f(\mathbf{W}_{\Pi_{s+\pi_{m+1-j}}})ds,$$

in $L^2(\Omega)$, as N tends to infinity, so that, from equation (3.26), Step 7, and the above, we conclude that

$$R(P^N, \Pi) \to \frac{1}{2}\sum_{j=1}^{m+1}\int_{\pi-\pi_{(m+1-j)}}^{t-\pi_{(m+1-j)}} \partial_{jj}^2 f(\mathbf{W}_{\Pi_{s+\pi_{(m+1-j)}}})ds, \tag{3.29}$$

in $L^2(\Omega)$-norm, as N tends to infinity.

Step 9. Finally, substituting (3.21) and (3.29) into (3.10), we obtain

$$\begin{aligned} f(\mathbf{W}_{\Pi_t}) - f(\mathbf{W}_{\Pi_\pi}) &= \sum_{j=1}^{m+1}\int_0^\infty \partial_j f(\mathbf{W}_{\Pi_{s+\pi_{m-j+1}}})1_{[\pi-\pi_{m+1-j}, t-\pi_{m+1-j}]}(s)\delta W_s \\ &+ \sum_{k>j}\int_{\pi-\pi_{m+1-j}}^{t-\pi_{m+1-j}} \partial_{kj}^2 f(\mathbf{W}_{\Pi_{s+\pi_{m+1-j}}})ds \\ &+ \frac{1}{2}\sum_{j=1}^{m+1}\int_{\pi-\pi_{(m+1-j)}}^{t-\pi_{(m+1-j)}} \partial_{jj}^2 f(\mathbf{W}_{\Pi_{s+\pi_{(m+1-j)}}})ds, \qquad a.s., \end{aligned}$$

which together with Schwarz' theorem, which guarantees that partial derivatives are symmetric, leads to the desired result. ☐

Corollary 3.4. *The formula given in Theorem 3.3 remains valid if* $f \in C^3(\mathbb{R}^{m+1})$.

Proof. Fix $t \geqslant \pi$, set $H = L^2([0,t])$, and define

$$\Omega_n = \left\{\sup_{s\leqslant t}|W_s| \leqslant n\right\}, \quad n \geqslant 1.$$

As f belongs to $C^3(\mathbb{R}^{m+1})$, there exists a sequence of functions $(f_n)_{n\geqslant 1}$ such that

$$f_n \in C_p^3(\mathbb{R}^{m+1}) \qquad \text{for all } n \geqslant 1,$$

and

$$f = f_n \quad \text{on } \left\{x = (x_0, \cdots, x_m) \in \mathbb{R}^{m+1} : |x_j| \leqslant n+1, \quad \text{for } j = 0, \cdots, m\right\}, \quad n \geqslant 1.$$

Now, realise that for any $j = 1, \cdots, m+1$ the sequence

$$\left(\Omega_n, \partial_j f_n(\boldsymbol{W}_{\Pi_{\cdot+\pi_{m-j+1}}}) \mathbb{1}_{(\pi-\pi_{m-j+1}, t-\pi_{m-j+1}]}\right)_{n\geqslant 1},$$

is a localising sequence for $\partial_j f(\boldsymbol{W}_{\Pi_{\cdot+\pi_{m-j+1}}}) \mathbb{1}_{(\pi-\pi_{m-j+1}, t-\pi_{m-j+1}]}$ as discussed in Section 2.4.3. Therefore, the Skorohod integral

$$\int_{\pi-\pi_{m-j+1}}^{t-\pi_{m-j+1}} \partial_j f(\boldsymbol{W}_{\Pi_{s+\pi_{m-j+1}}}) \delta W_s,$$

is well defined and equals

$$\int_{\pi-\pi_{m-j+1}}^{t-\pi_{m-j+1}} \partial_j f_n(\boldsymbol{W}_{\Pi_{s+\pi_{m-j+1}}}) \delta W_s, \quad \text{on } \Omega_n \text{ for all } n \geqslant 1.$$

Of course, $f_n \in C_p^3(\mathbb{R}^{m+1})$, and hence, by Theorem 3.3

$$\begin{aligned} f_n(\boldsymbol{W}_{\Pi_t}) = f_n(\boldsymbol{W}_{\Pi_\pi}) &+ \sum_{j=1}^{m+1} \int_{\pi-\pi_{(m+1-j)}}^{t-\pi_{(m+1-j)}} \partial_j f_n(\boldsymbol{W}_{\Pi_{s+\pi_{(m+1-j)}}}) \delta W_s \\ &+ \frac{1}{2} \sum_{j=1}^{m+1} \sum_{k=1}^{m+1} \int_\pi^t \partial_{jk}^2 f_n(\boldsymbol{W}_{\Pi_s}) ds, \qquad \text{a.s.}, \end{aligned}$$

for all $n \geqslant 1$. Using the equality of the above Skorohod integrals on Ω_n in the last formula gives

$$\begin{aligned} f(\boldsymbol{W}_{\Pi_t}) = f(\boldsymbol{W}_{\Pi_\pi}) &+ \sum_{j=1}^{m+1} \int_{\pi-\pi_{(m+1-j)}}^{t-\pi_{(m+1-j)}} \partial_j f(\boldsymbol{W}_{\Pi_{s+\pi_{(m+1-j)}}}) \delta W_s \\ &+ \frac{1}{2} \sum_{j=1}^{m+1} \sum_{k=1}^{m+1} \int_\pi^t \partial_{jk}^2 f(\boldsymbol{W}_{\Pi_s}) ds, \qquad \text{a.s. on } \Omega_n, \end{aligned}$$

for all $n \geqslant 1$, proving the corollary because $\Omega_n \uparrow \Omega$, as $n \to \infty$. □

3.3.2 The Diffusion Case

In this section, we present a generalisation of Theorem 3.3 to the case of diffusions which are solutions to equation (3.1), i.e.;

$$\begin{cases} dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t, \\ X_0 \ \ \text{given.} \end{cases}$$

Observe that if X is a solution to equation (3.1), then it admits the following decomposition

$$X_t = \begin{cases} \eta(\pi) + \int_\pi^t \mu(s)dW_s + \int_\pi^t \nu(s)ds, & \pi \leqslant t \leqslant T; \\ \eta(t), & 0 \leqslant t \leqslant \pi, \end{cases}$$

where η, μ and ν are some stochastic processes. But, if X is a continuous process with such decomposition then the following theorem holds true.

Theorem 3.5. *(c.f. Theorem 2.1 in Hu, Mohammed, and Yan (2004)) If* $(\eta(t))_{t\in[0,\pi]}$ *is deterministic and of bounded variation,* $\mu \in \mathbb{L}^{2,4}_{loc}$, *and* $\nu \in \mathbb{L}^{1,4}_{loc}$, *then for any* $f \in C^2(\mathbb{R}^{m+1})$

$$\begin{aligned} f(\mathbf{X}_{\Pi_t}) = f(\mathbf{X}_{\Pi_\pi}) &+ \sum_{j=1}^{m+1} \int_{\pi-\pi_{m-j+1}}^{t-\pi_{m-j+1}} \partial_j f(\mathbf{X}_{\Pi_{s+\pi_{m-j+1}}})\mu(s)\delta W_s \\ &+ \sum_{j=1}^{m+1} \int_\pi^t \partial_j f(\mathbf{X}_{\Pi_s})\nu(s - \pi_{m-j+1})ds \\ &+ \frac{1}{2}\sum_{j=1}^{m+1}\sum_{k=1}^{m+1} \int_\pi^t \partial^2_{jk} f(\mathbf{X}_{\Pi_s})\mu(s-\pi_{m-j+1})D_{s-\pi_{m-j+1}}X_{s-\pi_{m-k+1}}ds, \end{aligned}$$

almost surely.

Remark 3.3. The processes μ and ν in the above theorem are not necessarily adapted, and hence X considered by Hu, Mohammed, and Yan (2004) is less regular than our X being a solution to equation (3.1). The price is that Hu, Mohammed, and Yan require μ to be twice Malliavin differentiable. In our case, $\mu(s) = \sigma(s, X_s)$ with σ being Lipschitz continuous in the space variable, so that our μ is only once Malliavin differentiable. But, as our μ is an adapted process, we are still able to prove Itô's formula under slightly weaker conditions.

First, the following Lemma (a more general version can be found as Theorem 2.2.1. in Nualart (2013) where the coefficients of the diffusion are globally Lipschitz and with at most linear growth) states that our assumptions are sufficient to guarantee that X_t belongs to $\mathbb{D}^{1,p}$ for all $t \geqslant 0$, for some $p \geqslant 2$, once the initial condition has enough integrability. Here, and in what follows, we set $H = L^2([0,T])$ to be the Hilbert space associated with the Malliavin derivative

Lemma 3.6. *Fix* $p \geqslant 2$. *Let* $X = \{X_t : 0 \leqslant t \leqslant T\}$ *be the solution to* (3.1) *where* $X_0 \in L^p(\mathbb{P})$ *and the coefficients satisfy assumptions 3.1 and 3.2. Then,* X_t *belongs to* $\mathbb{D}^{1,p}$ *for any* $0 \leqslant t \leqslant T$, *satisfies*

$$\sup_{0 \leqslant r \leqslant t} \mathbb{E}\left[\sup_{r \leqslant s \leqslant T} |D_r X_s|^p\right] < \infty, \tag{3.30}$$

and the derivative process $D_r X_s$ *satisfies the following linear SDE*

$$D_r X_t = \sigma(r, X_r) + \int_r^t \tilde{\sigma}_s D_r X_s \, dW_s + \int_r^t \tilde{b}_s D_r X_s \, ds,$$

for $r \leqslant t$, *a.s., and*

$$D_r X_t = 0,$$

for $r > t$, *a.s.; where* $\tilde{\sigma}$ *and* $\tilde{b}$ *are uniformly bounded and adapted processes.*

Proof. It follows from Theorem 2.2.1 in Nualart (2013). □

Our main result states as follows. The general structure and ideas in the proof of Theorem 3.3 remain useful to achieve the proof of the more general result. That is, we will start by setting $(P^N)_{N \geqslant 1}$ a sequence of partitions of the interval $[\pi, t]$, and writing

$f(\mathbf{X}_{\Pi_t})$ as the result of a telescopic sum of f evaluated in the points of the partition P^N. Then, we will use a Taylor expansion of second order and investigate the convergence of terms analogous to the ones in the right hand side of equation (3.10). However, it is worth mentioning that some difficulties arise at passing to the diffusion case. First of all, terms (3.11) and (3.22) and are more complex due to the fact that instead of having differences of the form $(W_{s_i - \pi_{m-(j-1)}} - W_{s_{i-1} - \pi_{m-(j-1)}})$ we have to deal with the difference of the diffusion, which involves the drift and also the diffusion coefficient σ integrated agains the Brownian motion. As a result, the terms in (3.12) no longer belong to the class of smooth elementary process and therefore we cannot use Proposition 2.7 to calculate its Skorohod integral as we did before. The same happens to the terms in (3.24).

Theorem 3.7. *(Itô Type Formula for Multiple Delay: Diffusion Case) Suppose that* $X_0 \in L^4(\mathbb{P})$ *and that assumptions 3.1 and 3.2 hold. Let* X *be the unique strong solution to equation (3.1), and* $\Pi = \{0 = \pi_0 < \cdots < \pi_m = \pi\}$ *a collection of delay points. Consider the delay vector process* $\mathbf{X}_\Pi$*, and a function* $f : \mathbb{R}^{m+1} \to \mathbb{R}$*. If* f *belongs to* $C^2(\mathbb{R}^{m+1})$*, then the process* $\{f(\mathbf{X}_{\Pi_t}) : \pi \leqslant t \leqslant T\}$ *satisfies*

$$\begin{aligned} f(\mathbf{X}_{\Pi_t}) =& f(\mathbf{X}_{\Pi_\pi}) + \sum_{j=1}^{m+1} \int_{\pi - \pi_{(m-j+1)}}^{t - \pi_{(m-j+1)}} \partial_j f(\mathbf{X}_{\Pi_{s+\pi_{(m-j+1)}}}) \sigma(s, X_s) \delta W_s \\ &+ \sum_{j=1}^{m+1} \int_\pi^t \partial_j f(\mathbf{X}_{\Pi_s}) b(s - \pi_{m-j+1}, X_{s-\pi_{m-j+1}}) ds \\ &+ \frac{1}{2} \sum_{j=1}^{m+1} \int_\pi^t \partial_{jj}^2 f(\mathbf{X}_{\Pi_s}) \sigma^2(s - \pi_{m-j+1}, X_{s-\pi_{m-j+1}}) ds \\ &+ \sum_{k>j} \int_\pi^t \partial_{jk}^2 f(\mathbf{X}_{\Pi_s}) \sigma(s - \pi_{m-j+1}, X_{s-\pi_{m-j+1}}) D_{s-\pi_{m-j+1}} X_{s-\pi_{m-k+1}} ds, \qquad a.s., \end{aligned}$$

where

$$\int_{\pi - \pi_{(m-j+1)}}^{t - \pi_{(m-j+1)}} \partial_j f(\mathbf{X}_{\Pi_{s+\pi_{(m-j+1)}}}) \sigma(s, X_s) \delta W_s, \qquad j = 1, \cdots, m+1,$$

are stochastic integrals in the Skorohod sense.

Proof. Let $t \in [\pi, T]$ be a fixed time and

$$P^N = \{\pi = s_0^N < s_1^N < \cdots < s_N^N = t\}, \qquad N \geqslant 1,$$

be a sequence of partitions of the interval $[\pi, t]$ such that $|P^N|$ tends to zero as N goes to infinity. Besides, let us assume that

$$|P^N| = \max_i\{s_{i+1}^N - s_i^N\} \leqslant \pi, \qquad \text{for all} \quad N \geqslant 1. \tag{3.31}$$

We can make this assumption without lost of generality since the mesh tends to zero as N goes to infinity. Hereafter, we will write s_i instead of s_i^N for notational convenience. Furthermore, by a localisation argument as the one in the proof of Corollary 3.4 we may assume that $f \in C_b^2(\mathbb{R}^{m+1})$.

For any $N \geqslant 1$, using Taylor's expansion, we obtain

$$\begin{aligned} f(\mathbf{X}_{\Pi_t}) - f(\mathbf{X}_{\Pi_\pi}) &= \sum_{i=1}^N [f(\mathbf{X}_{\Pi_{s_i}}) - f(\mathbf{X}_{\Pi_{s_{i-1}}})] \\ &= \sum_{i=1}^N \left\{ Df(\mathbf{X}_{\Pi_{s_{i-1}}})(\mathbf{X}_{\Pi_{s_i}} - \mathbf{X}_{\Pi_{s_{i-1}}}) + \frac{1}{2}(\mathbf{X}_{\Pi_{s_i}} - \mathbf{X}_{\Pi_{s_{i-1}}})^\intercal D^{(2)} f(\xi_{\Pi_i})(\mathbf{X}_{\Pi_{s_i}} - \mathbf{X}_{\Pi_{s_{i-1}}}) \right\}, \end{aligned}$$

where $\xi_{\Pi_i} = \theta \mathbf{X}_{\Pi_{s_{i-1}}} + (1-\theta)\mathbf{X}_{\Pi_{s_i}}$, for some random θ between zero and one. Thus

$$f(\mathbf{X}_{\Pi_t}) - f(\mathbf{X}_{\Pi_\pi}) = S(P^N, \Pi) + R(P^N, \Pi), \tag{3.32}$$

where

$$S(P^N, \Pi) = \sum_{j=1}^{m+1} \sum_{i=1}^N \partial_j f(\mathbf{X}_{\Pi_{s_{i-1}}})(X_{s_i - \pi_{m-(j-1)}} - X_{s_{i-1} - \pi_{m-(j-1)}}),$$

and

$$R(P^N, \Pi) = \frac{1}{2} \sum_{j,k=1}^{m+1} \sum_{i=1}^N (X_{s_i - \pi_{m-(j-1)}} - X_{s_{i-1} - \pi_{m-(j-1)}}) \partial_{jk}^2 f(\xi_{\Pi_i})(X_{s_i - \pi_{m-(k-1)}} - X_{s_{i-1} - \pi_{m-(k-1)}}).$$

Next, we investigate the convergence of these sums as N tends to infinity. More precisely, we are going to prove that both terms converge in probability as N tends to infinity. The proof will be decomposed into several steps to make it clearer and easier to read.

Step 1. First notice that

$$\begin{aligned} X_{s_i-\pi_{m-(j-1)}} - X_{s_{i-1}-\pi_{m-(j-1)}} &= \int_{s_{i-1}-\pi_{m-(j-1)}}^{s_i-\pi_{m-(j-1)}} b(s, X_s)ds \\ &+ \int_{s_{i-1}-\pi_{m-(j-1)}}^{s_i-\pi_{m-(j-1)}} \sigma(s, X_s)dW_s, \qquad \text{a.s.,} \end{aligned} \tag{3.33}$$

since X satisfies equation (3.1). So, we can write

$$S(P^N, \Pi) = S_b(P^N, \Pi) + S_\sigma(P^N, \Pi), \qquad \text{a.s.,} \tag{3.34}$$

where

$$S_b(P^N, \Pi) = \sum_{j=1}^{m+1} \sum_{i=1}^{N} \partial_j f(\mathbf{X}_{\Pi_{s_{i-1}}}) \int_{s_{i-1}-\pi_{m-(j-1)}}^{s_i-\pi_{m-(j-1)}} b(s, X_s)ds, \tag{3.35}$$

and

$$S_\sigma(P^N, \Pi) = \sum_{j=1}^{m+1} \sum_{i=1}^{N} \partial_j f(\mathbf{X}_{\Pi_{s_{i-1}}}) \int_{s_{i-1}-\pi_{m-(j-1)}}^{s_i-\pi_{m-(j-1)}} \sigma(s, X_s)dW_s. \tag{3.36}$$

Hence, it is sufficient to study these two terms as N tends to infinity to deduce the behaviour of $S(P^N, \Pi)$.

Step 2. Using the linearity of the Riemman integral, assumption 3.2 , and the fact that f has bounded derivatives of first order, we obtain

$$\sum_{i=1}^{N} \int_{s_{i-1}-\pi_{m-(j-1)}}^{s_i-\pi_{m-(j-1)}} \partial_j f(\mathbf{X}_{\Pi_{s_{i-1}}})b(s, X_s)ds \to \int_{\pi-\pi_{m-(j-1)}}^{t-\pi_{m-(j-1)}} \partial_j f(\mathbf{X}_{\Pi_{s+\pi_{(m-j+1)}}})b(s, X_s)ds,$$

in probability, as N tends to infinity, for each $j = 1, \cdots, m$. Thus

$$S_b(P^N, \Pi) \to \sum_{j=1}^{m+1} \int_{\pi - \pi_{m-(j-1)}}^{t - \pi_{m-(j-1)}} \partial_j f(\mathbf{X}_{\Pi_{s+\pi_{(m-j+1)}}}) b(s, X_s) ds, \tag{3.37}$$

in probability, as N tends to infinity.

Step 3. To study $S_\sigma(P^N, \Pi)$, first notice that

$$\begin{aligned}\int_{s_{i-1} - \pi_{m-(j-1)}}^{s_i - \pi_{m-(j-1)}} \sigma(s, X_s) dW_s &= \int_0^T \sigma(s, X_s) 1_{A_i^j}(s) dW_s \\ &= \int_0^T \sigma(s, X_s) 1_{A_i^j}(s) \delta W_s, \qquad \text{a.s.,}\end{aligned}$$

where A_i^j stands for the interval $(s_{i-1} - \pi_{m-(j-1)}, s_i - \pi_{m-(j-1)}]$, and the last equality follows due to the fact that the Itô integral coincides with the Skorohod integral for adapted square integrable processes (See Proposition 1.3.7 in Nualart, 2013). Thus, we can write (3.36) as

$$S_\sigma(P^N, \Pi) = \sum_{j=1}^{m+1} \sum_{i=1}^{N} \partial_j f(\mathbf{X}_{\Pi_{s_{i-1}}}) \delta(\sigma(\cdot, X_\cdot) 1_{A_i^j}(\cdot)),$$

almost surely.

Now, we are going to introduce some ancillary quantities. Let

$$F_i^j = \partial_j f(\mathbf{X}_{\Pi_{s_{i-1}}}), \qquad j = 1, \cdots, m+1,$$

and

$$v_i^j(s) = \sigma(s, X_s) 1_{A_i^j}(s), \quad s \leqslant T,$$

so that

$$S_\sigma(P^N, \Pi) = \sum_{j=1}^{m+1} S_{\sigma_j}(P^N, \Pi), \qquad \text{a.s.,} \tag{3.38}$$

where

$$S_{\sigma_j}(P^N,\Pi)=\sum_{i=1}^{N}F_i^j\delta(\nu_i^j).$$

In the next three steps we will find the limit of $S_\sigma(P^N,\Pi)$ as $N\to\infty$.

Step 4. First, we are going to study the terms

$$S_{\sigma_j}(P^N,\Pi)=\sum_{i=1}^{N}F_i^j\delta(\nu_i^j),\qquad j=1,\cdots,m+1. \tag{3.39}$$

The idea is to use Proposition 2.7 to write each $S_{\sigma_j}(P^N,\Pi)$ in a way that we can deduce its behaviour as N tends to infinity.

First, notice that Lemma 2.3 and Lemma 3.6 together imply that the random variables F_i^j belong to $\mathbb{D}^{1,2}$ since f is continuously differentiable with bounded partial derivatives of second order by localisation. Moreover, we have

$$D(F_i^j)=\sum_{k=1}^{m+1}\partial^2_{kj}f(\mathbf{X}_{\Pi_{s_{i-1}}})D(X_{s_{i-1}-\pi_{m-(k-1)}}), \tag{3.40}$$

almost surely.

As explained above, the terms ν_i^j belong to $\mathrm{Dom}(\delta)$. Finally, for all $s\leqslant T$,

$$\begin{aligned}\mathbb{E}\left[|F_i^j\nu_i^j(s)|^2\right]&=\mathbb{E}\left[|\partial_j f(\mathbf{X}_{\Pi_{s_{i-1}}})\sigma(s,X_s)\mathbb{1}_{A_i^j}(s)|^2\right]\\&\leqslant C\mathbb{E}\left[|\sigma(s,X_s)|^2\right]\\&\leqslant 2CK^2\mathbb{E}[|X_s|^2]+2CM^2,\end{aligned}$$

by using the assumption that we made about first order partial derivatives of f being bounded, followed by assumptions 3.1 and 3.2 . So,

$$\begin{aligned}\int_0^T \mathbb{E}\left[|F_i^j v_i^j(s)|^2\right] ds &\leqslant 2CK^2 \int_0^T \mathbb{E}[|X_s|^2] ds + 2CM^2T \\ &\leqslant 2CK^2 \int_0^T \mathbb{E}\left[\sup_{0\leqslant s\leqslant T} |X_s|^2\right] ds + 2CM^2T \\ &\leqslant 2CT(K^2C_0 + M^2),\end{aligned}$$

where the last inequality follows by (3.2) and the constant C_0 depends only on T, K, M, and $\mathbb{E}[|X_0|^2]$. Consequently,

$$\left(\int_0^T \mathbb{E}\left[|F_i^j v_i^j(s)|^2\right] ds\right)^{1/2} < \infty.$$

Thus, we have F_i^j in $\mathbb{D}^{1,2}$, and v_i^j in $\mathrm{Dom}(\delta)$ such that $F_i^j v_i^j$ belongs to $L^2(\Omega; H)$. Then, Proposition 2.7 implies that $F_i^j v_i^j(s)$ also belongs to $\mathrm{Dom}(\delta)$. Moreover,

$$\delta(F_i^j v_i^j) = F_i^j \delta(v_i^j) - \int_0^T D_s F_i^j v_i^j(s) ds, \qquad \text{a.s.,} \tag{3.41}$$

since

$$\begin{aligned}F_i^j \delta(v_i^j) - \int_0^T D_s F_i^j v_i^j(s) ds = \partial_j f(\mathbf{X}_{\Pi_{s_{i-1}}}) \int_0^T \sigma(s, X_s) 1_{A_i^j}(s) dW_s \\ - \int_0^T \sum_{k=1}^{m+1} \partial_{kj}^2 f(\mathbf{X}_{\Pi_{s_{i-1}}}) D_s(X_{s_{i-1}-\pi_{m-(k-1)}}) \sigma(s, X_s) 1_{A_i^j}(s) ds,\end{aligned}$$

is square integrable, thanks to assumptions 3.1 and 3.2, f having bounded first and second order partial derivatives by localisation, and the fact that X_t belongs to $\mathbb{D}^{1,2}$ and

$$\sup_{0\leqslant r\leqslant t} \mathbb{E}\left[\sup_{r\leqslant s\leqslant T} |D_r X_s|^2\right] < \infty,$$

for any $0 \leqslant t \leqslant T$, by Lemma 3.6.

Reordering the terms in equation (3.41) we obtain

$$F_i^j\delta(v_i^j) = \delta(F_i^j v_i^j) + \int_0^T D_s F_i^j v_i^j(s)ds, \qquad \text{a.s.},$$

which holds for each $i = 1, \cdots, N$ and $j = 1, \cdots, m+1$, and consequently

$$\begin{aligned}\sum_{i=1}^N F_i^j\delta(v_i^j) &= \sum_{i=1}^N \delta(F_i^j v_i^j) + \sum_{i=1}^N \int_0^T D_s F_i^j v_i^j(s)ds \\ &= \delta\left(\sum_{i=1}^N F_i^j v_i^j\right) + \int_0^T \sum_{i=1}^N D_s F_i^j v_i^j(s)ds, \qquad \text{a.s.},\end{aligned}$$

where the last inequality follows by the linearity of both Skorohod and Lebesgue integrals.

Substituting the last expression into (3.39), we obtain

$$S_{\sigma_j}(P^N, \Pi) = \delta\left(\sum_{i=1}^N F_i^j v_i^j\right) + \int_0^T \sum_{i=1}^N D_s F_i^j v_i^j(s)ds, \quad \text{a.s.}, \tag{3.42}$$

for all $j = 1, \cdots, m+1$. So, it suffices to investigate the convergence of the terms on the right hand side.

Step 5. To study the convergence of the first term on the right hand side of equation (3.42) let us introduce the following ancillary quantities:

$$u_j^N(s) = \sum_{i=1}^N F_i^j(s)v_i^j(s), \qquad N \geqslant 1,$$

and

$$u_j^\infty = F^j v^j, \qquad j = 1, \cdots, m+1,$$

where

$$F^j(s) = \partial_j f(\mathbf{X}_{\Pi_{s+\pi_{m-(j-1)}}}),$$

and

$$v^j(s) = \sigma(s, X_s)1_{[\pi-\pi_{m-(j-1)}, t-\pi_{m-(j-1)}]}(s).$$

Next, we are going to prove that

$$u_j^N \to u_j^\infty, \qquad \text{in } \mathbb{D}^{1,2}(H), \quad \text{as } N \to \infty. \tag{3.43}$$

So, we fix $1 \leqslant j \leqslant m+1$ and estimate

$$\mathbb{E}\left[\int_0^T |u_j^\infty(s) - u_j^N(s)|^2 ds\right] + \mathbb{E}\left[\int_0^T \int_0^T |D_r u_j^\infty(s) - D_r u_j^N(s)|^2 dr ds\right]. \tag{3.44}$$

Using the definitions of u_j^∞ and u_j^N, the first summand reads

$$\begin{aligned}
&\mathbb{E}\left[\int_0^T |u_j^\infty(s) - u_j^N(s)|^2 ds\right] \\
&= \sum_{i=1}^N \int_{s_{i-1}}^{s_i} \mathbb{E}\left[|\partial_j f(\mathbf{X}_{\Pi_s}) - \partial_j f(\mathbf{X}_{\Pi_{s_{i-1}}})|^2 |\sigma(s - \pi_{m-j+1}, X_{s-\pi_{m-j+1}})|^2\right] ds.
\end{aligned}$$

So, by using mean value theorem followed by Cauchy-Schwarz inequality, we obtain

$$\mathbb{E}\left[\int_0^T |u_j^\infty(s) - u_j^N(s)|^2 ds\right] \leqslant C \sum_{i=1}^N \sum_{l=1}^{m+1} \left\{ \left(I_{il}^{(1)}\right)^{1/2} \left(I_{il}^{(2)}\right)^{1/2} \right\}, \tag{3.45}$$

for some constant $C > 0$, and

$$I_{il}^{(1)} = \int_{s_{i-1}}^{s_i} \mathbb{E}[|\partial_{jl}^2 f(\theta \mathbf{X}_{\Pi_s} + (1-\theta)\mathbf{X}_{\Pi_{s_{i-1}}})|^4 |\sigma(s - \pi_{m-j+1}, X_{s-\pi_{m-j+1}})|^4] ds,$$

where $\theta \in [0,1]$ is a random number; and

$$I_{il}^{(2)} = \int_{s_{i-1}}^{s_i} \mathbb{E}[|X_{s-\pi_{m-l+1}} - X_{s_{i-1}-\pi_{m-l+1}}|^4] ds.$$

Now, notice that for all $i = 1, \cdots, N$, and $l = 1, \cdots, m+1$

$$I_{il}^{(1)} \leqslant C \int_{s_{i-1}}^{s_i} \mathbb{E}[(K|X_{s-\pi_{m-j+1}}| + M)^4] ds,$$

since all the second order partial derivatives of f are bounded by localisation, and σ satisfies assumptions 3.1 and 3.2. Moreover, using equation (3.2) we obtain

$$I_{il}^{(1)} \leqslant C_0(s_i - s_{i-1}), \tag{3.46}$$

where C_0 is a constant depending on T, K, M and $\mathbb{E}[|X_0|^4]$.

On the other hand, for all $l = 1, \cdots, m+1$, we have

$$\begin{aligned}
&\mathbb{E}\left[(X_{s-\pi_{m-l+1}} - X_{s_{i-1}-\pi_{m-l+1}})^4\right] \\
&= \mathbb{E}\left[\left(\int_{s_{i-1}-\pi_{m-l+1}}^{s-\pi_{m-l+1}} b(u, X_u)du + \int_{s_{i-1}-\pi_{m-l+1}}^{s-\pi_{m-l+1}} \sigma(u, X_u)dW_u\right)^4\right] \\
&\leqslant \mathbb{E}\left[\left(2\left(\int_{s_{i-1}-\pi_{m-l+1}}^{s-\pi_{m-l+1}} b(u, X_u)du\right)^2 + 2\left(\int_{s_{i-1}-\pi_{m-l+1}}^{s-\pi_{m-l+1}} \sigma(u, X_u)dW_u\right)^2\right)^2\right] \\
&\leqslant 8\mathbb{E}\left[\left(\int_{s_{i-1}-\pi_{m-l+1}}^{s-\pi_{m-l+1}} b(u, X_u)du\right)^4\right] + 8\mathbb{E}\left[\left(\int_{s_{i-1}-\pi_{m-l+1}}^{s-\pi_{m-l+1}} \sigma(u, X_u)dW_u\right)^4\right],
\end{aligned}$$

and therefore

$$\begin{aligned}
I_{il}^{(2)} \leqslant 8\int_{s_{i-1}}^{s_i} &\mathbb{E}\left[\left(\int_{s_{i-1}-\pi_{m-l+1}}^{s-\pi_{m-l+1}} b(u, X_u)du\right)^4\right] ds \\
&+ 8\int_{s_{i-1}}^{s_i} \mathbb{E}\left[\left(\int_{s_{i-1}-\pi_{m-l+1}}^{s-\pi_{m-l+1}} \sigma(u, X_u)dW_u\right)^4\right] ds \\
&\leqslant 8C_0^4 \frac{(s_i - s_{i-1})^5}{5} + 8\int_{s_{i-1}}^{s_i} C_2\mathbb{E}\left[\left(\int_{s_{i-1}-\pi_{m-l+1}}^{s_i-\pi_{m-l+1}} \sigma^2(u, X_u)du\right)^2\right] ds \\
&\leqslant 8C_0^4 \frac{(s_i - s_{i-1})^5}{5} + 8C_2\frac{(s_i - s_{i-1})^5}{5}(s_i - s_{i-1}),
\end{aligned}$$

where the constant C_2 comes from the Burkholder-Davis-Gundy inequality.

Consequently, for all $l = 1, \cdots, m+1$, we have

$$I_{il}^{(2)} \leqslant C\frac{(s_i - s_{i-1})^5}{5}, \tag{3.47}$$

where the constant C is a constant which depends on C_0, C_2, m, Π and $\mathbb{E}[|X_0|^4]$.

So, putting together equations (3.45) to (3.47) we obtain the following inequality

$$\mathbb{E}\left[\int_0^T |u_j^\infty(s) - u_j^N(s)|^2 ds\right] \leqslant C \sum_{i=1}^N \sum_{l=1}^{m+1} (s_i - s_{i-1})^{1/2} \frac{(s_i - s_{i-1})^{5/2}}{\sqrt{5}},$$

that is, the first summand of equation (3.44) is bounded by

$$C(m+1) \sum_{i=1}^N (s_i - s_{i-1})^{1/2} (s_i - s_{i-1})^{5/2},$$

which converges to zero when $N \to \infty$ as

$$\sum_{i=1}^N (s_i - s_{i-1}) = t - \pi.$$

Now, working out $D_r u_j^\infty(s)$ and $D_r u_j^N(s)$ in the second sum of (3.44), we obtain

$$\begin{aligned}
&\mathbb{E}\left[\int_0^T \int_0^T |D_r u_j^\infty(s) - D_r u_j^N(s)|^2 drds\right] \\
&= \mathbb{E}\left[\int_0^T \int_0^T |D_r \left(\partial_j f(\mathbf{X}_{\Pi_{s+\pi_{m-(j-1)}}})\sigma(s, X_s) 1_{[\pi-\pi_{m-(j-1)}, t-\pi_{m-(j-1)}]}(s)\right) - D_r \left(\sum_{i=1}^N F_i^j(s) v_i^j(s)\right) |^2 drds\right] \\
&= \mathbb{E}\left[\int_0^T \int_0^T |D_r \left(\partial_j f(\mathbf{X}_{\Pi_{s+\pi_{m-(j-1)}}})\sigma(s, X_s) 1_{[\pi-\pi_{m-(j-1)}, t-\pi_{m-(j-1)}]}(s)\right)\right. \\
&\qquad\qquad \left. - D_r \left(\sum_{i=1}^N \partial_j f(\mathbf{X}_{\Pi_{s_{i-1}}})\sigma(s, X_s) 1_{A_i^j}(s)\right) |^2 drds\right] \\
&= \mathbb{E}\left[\int_0^T \int_0^T \left|\sum_{i=1}^N D_r \left(\partial_j f(\mathbf{X}_{\Pi_{s+\pi_{m-(j-1)}}})\sigma(s, X_s) 1_{A_i^j}(s)\right)\right.\right. \\
&\qquad\qquad \left.\left. - \sum_{i=1}^N D_r \left(\partial_j f(\mathbf{X}_{\Pi_{s_{i-1}}})\sigma(s, X_s) 1_{A_i^j}(s)\right)\right|^2 drds\right],
\end{aligned}$$

By using the product rule followed by calculations similar to the previous one, and dominated convergence we also can conclude that this term convergest to zero as N goes to infinity.

All in all, both summands of (3.44) converge to zero when $N \to \infty$ which finally proves (3.43). Therefore, using Proposition 2.6, we can conclude that

$$\delta(u_j^N) \to \delta(u_j^\infty), \quad \text{in } L^2(\Omega) \text{ as } N \to \infty,$$

for $j = 1, \cdots, m+1$. That is,

$$\delta\left(\sum_{i=1}^{N} F_i^j u_i^j\right) \to \delta\left(\partial_j f(\mathbf{X}_{\Pi_{s+\pi_{m-(j-1)}}})\sigma(s, X_s)1_{[\pi-\pi_{m-(j-1)}, t-\pi_{m-(j-1)}]}(s)\right), \tag{3.48}$$

in $L^2(\Omega)$, as N tends to infinity, for $j = 1, \cdots, m+1$.

Step 6. Now, for the last term in equation (3.42), i.e.,

$$\int_0^T \sum_{i=1}^{N} D_s F_i^j v_i^j(s) ds,$$

we claim that it converges in probability to

$$\sum_{k=j+1}^{m+1} \int_\pi^t \partial_{jk}^2 f(\mathbf{X}_{\Pi_s}) D_{s-\pi_{m-j+1}}(X_{s-\pi_{m-k+1}})\sigma(s - \pi_{m-j+1}, X_{s-\pi_{m-j+1}}) 1_{j \leqslant m} ds,$$

as $N \to \infty$.

Firstly, by using the definition of v_i^j, equation (3.40), and the linearity of the Rieman integral, we can write

$$\begin{aligned} \int_0^T \sum_{i=1}^{N} D_s F_i^j v_i^j(s) ds &= \int_0^T \sum_{i=1}^{N} \sum_{k=1}^{m+1} \partial_{kj}^2 f(\mathbf{X}_{\Pi_{s_{i-1}}}) D_s(X_{s_{i-1}-\pi_{m-(k-1)}})\sigma(s, X_s) 1_{A_i^j}(s) ds \\ &= \sum_{i=1}^{N} \sum_{k=1}^{m+1} \int_0^T \partial_{kj}^2 f(\mathbf{X}_{\Pi_{s_{i-1}}}) D_s(X_{s_{i-1}-\pi_{m-(k-1)}})\sigma(s, X_s) 1_{A_i^j}(s) ds. \end{aligned}$$

Now, the key point here is to notice that by Lemma 3.6 $D_s(X_{s_{i-1}-\pi_{m-(k-1)}})$ is equal to zero if $s > s_{i-1} - \pi_{m-k+1}$. So, we have

$$D_s(X_{s_{i-1}-\pi_{m-(k-1)}}) = 0, \qquad \text{if } \; s \in A_i^j, \text{a.s., for } j \geqslant k,$$

and consequently we can write

$$\begin{aligned}
&\int_0^T \sum_{i=1}^N D_s F_i^j v_i^j(s)ds \\
&= \sum_{i=1}^N \sum_{k=j+1}^{m+1} \int_0^T \partial_{kj}^2 f(\mathbf{X}_{\Pi_{s_{i-1}}}) D_s(X_{s_{i-1}-\pi_{m-(k-1)}})\sigma(s, X_s) 1_{A_i^j}(s)ds 1_{\{j \leqslant m\}} \\
&= \sum_{k=j+1}^{m+1} \sum_{i=1}^N \int_0^T \partial_{kj}^2 f(\mathbf{X}_{\Pi_{s_{i-1}}}) D_s(X_{s_{i-1}-\pi_{m-(k-1)}})\sigma(s, X_s) 1_{A_i^j}(s)ds 1_{\{j \leqslant m\}}, \qquad \text{a.s.}
\end{aligned}$$

But, for each fixed k and $j \leqslant m$ we have

$$\begin{aligned}
&\sum_{i=1}^N \int_0^T \partial_{kj}^2 f(\mathbf{X}_{\Pi_{s_{i-1}}}) D_s(X_{s_{i-1}-\pi_{m-(k-1)}})\sigma(s, X_s) 1_{A_i^j}(s)ds \\
&= \sum_{i=1}^N \int_{s_{i-1}-\pi_{m-(j-1)}}^{s_i-\pi_{m-(j-1)}} \partial_{kj}^2 f(\mathbf{X}_{\Pi_{s_{i-1}}}) D_s(X_{s_{i-1}-\pi_{m-(k-1)}})\sigma(s, X_s)ds \\
&= \sum_{i=1}^N \int_{s_{i-1}}^{s_i} \partial_{kj}^2 f(\mathbf{X}_{\Pi_{s_{i-1}}}) D_{s-\pi_{m-(j-1)}}(X_{s_{i-1}-\pi_{m-(k-1)}})\sigma(s - \pi_{m-(j-1)}, X_{s-\pi_{m-(j-1)}})ds,
\end{aligned}$$

which converges almost surely to

$$\int_\pi^t \partial_{kj}^2 f(\mathbf{X}_{\Pi_s}) D_{s-\pi_{m-(j-1)}}(X_{s-\pi_{m-(k-1)}})\sigma(s - \pi_{m-(j-1)}, X_{s-\pi_{m-(j-1)}})ds,$$

by domain convergence using that the second order partial derivatives of f are bounded by localisation, σ satisfies assumptions 3.1 and 3.2, $X_0 \in L^4(\mathbb{P})$, and Lemma 3.6.

Therefore, we have that

$$\int_0^T \sum_{i=1}^N D_s F_i^j v_i^j(s)ds \to \sum_{k=j+1}^{m+1} \int_{\pi-\pi_{m-j+1}}^{t-\pi_{m-j+1}} \partial_{kj}^2 f(\mathbf{X}_{\Pi_{s+\pi_{m-j+1}}}) D_s(X_{s-\pi_{m-(k-1)}})\sigma(s, X_s)ds \cdot 1_{\{j \leqslant m\}}, \tag{3.49}$$

in $L^2(\Omega)$, as N tends to infinity.

Step 7. Finally,

$$R(P^N, \Pi) \to \frac{1}{2} \sum_{j=1}^{m+1} \int_{\pi - \pi_{(m-j+1)}}^{s - \pi_{(m-j+1)}} \partial_{jj}^2 f(\mathbf{X}_{\Pi_{s+\pi_{(m-j+1)}}}) \sigma^2(s, X_s) ds, \tag{3.50}$$

in probability, as $N \to \infty$, can be shown in the same way as in the proof of Proposition 2.3 in Hu, Mohammed, and Yan (2004). We could have used this proposition when proving our Ito's formula for delay-functionals of Brownian motion, but we gave a different proof based on the stronger assumption of f being three times differentiable.

Step 8. To conclude, notice that taking the limit, as N tends to infinity, in equation (3.32) and taking into account the previous steps we obtain

$$\begin{aligned} f(\mathbf{X}_{\Pi_t}) - f(\mathbf{X}_{\Pi_\pi}) &= \sum_{j=1}^{m+1} \int_{\pi - \pi_{(m-j+1)}}^{t - \pi_{(m-j+1)}} \partial_j f(\mathbf{X}_{\Pi_{s+\pi_{(m-j+1)}}}) \sigma(s, X_s) \delta W_s \\ &+ \sum_{j=1}^{m+1} \int_\pi^t \partial_j f(\mathbf{X}_{\Pi_s}) b(s - \pi_{m-j+1}, X_{s-\pi_{m-j+1}}) ds \\ &+ \frac{1}{2} \sum_{j=1}^{m+1} \int_\pi^t \partial_{jj}^2 f(\mathbf{X}_{\Pi_s}) \sigma^2(s - \pi_{m-j+1}, X_{s-\pi_{m-j+1}}) ds \\ &+ \sum_{k>j} \int_\pi^t \partial_{jk}^2 f(\mathbf{X}_{\Pi_s}) D_{s-\pi_{m-j+1}} X_{s-\pi_{m-k+1}} \sigma(s - \pi_{m-j+1}, X_{s-\pi_{m-j+1}}) ds, \end{aligned}$$

almost surely, and the result follows. □

4

A Class of Non-Markovian Nonlinear Stochastic Processes

Then we asked ourselves: Can you define all diffusion processes by just martingale properties? It looked like it unified different points of view: Kolmogorov and Feller through the PDE have one point of view, stochastic differential people have another point of view, semigroup theory has still another point of view. But the martingale point of view unifies them.

— **S. R. Srinivasa Varadhan**

Notices Amer. Math. Soc. (2008)

THE aim of this chapter is to study a class of $\mathbb{R}$-valued stochastic processes associated with families of second order differential operators of the form $\mathcal{A} = \{A[\mu] : \mu \in \mathcal{P}(\mathbb{R}^{(m+1)})\}$ with diffusion and drift coefficients given by $\sigma = \sigma(x, \mu)$, and $b = b(x, \mu)$, respectively. To formalise this idea we introduce a nonlinear martingale problem. Existence of solutions to such problem under mild assumptions on the coefficients is proved by using an approximation scheme.

4.1 Introduction

The martingale problem associated with a class of Levy generators

$$A_t = L_t + K_t, \qquad t \geqslant 0, \tag{4.1}$$

where

$$L_t f(x) = \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d} a_{ij}(t,x)\frac{\partial^2}{\partial x_i \partial x_j} f(x) + \sum_{i=1}^{d} b_i(t,x)\frac{\partial}{\partial x_i} f(x), \qquad t \geqslant 0,$$

and

$$K_t f(x) = \int \left(f(x+y) - f(x) - \frac{y \cdot \nabla f(x)}{1+|y|^2} \right) M(t,x;dy),$$

was studied by Stroock (1975) as an extension of the martingale problem, corresponding to diffusion processes, introduced by Stroock and Varadhan in Stroock and Varadhan (1969a); Stroock and Varadhan (1969c). More precisely, the martingale problem set by Stroock consists of finding for each (s,x) a probability measure $\mathbb{P}$ defined on the space of right-continuous functions having left limits such that

$$\mathbb{P}(X_s = x) = 1,$$

and the process

$$M_t = f(X_t) - \int_s^t \mathcal{L}_u f(X_u)du, \qquad t \geqslant s,$$

is a $\mathbb{P}$-martingale for all $f \in C_b^\infty(\mathbb{R}^d)$. Here, X represents the canonical process, i.e. $X_t(\omega) = \omega(t)$. Stroock proved existence of solutions under continuity assumptions on the coefficients. In addition, he demonstrated that uniqueness holds for various special cases. Since then, the martingale formulation has proven to be very useful because of the level of generality –in terms of the coefficients– that it allows to treat. In particular, the martingale approach was extended to cover nonlinear, in the sense of McKean, stochastic processes (see e.g. Sznitman, 1991 and Méléard, 1996); and more recently to cover other kinds of nonlinearity (see Kolokoltsov, 2010).

In this chapter, we will use the martingale approach to investigate the existence of a class of stochastic processes determined by families of operators (with common domain $\mathcal{D}$) of the form

$$\mathcal{A} = \{A[\mu] : \mathcal{D} \subset C^2(\mathbb{R}) \to C(\mathbb{R}), \quad \mu \in \mathcal{P}(\mathbb{R}^{(m+1)})\},$$

given by

$$A[\mu]f(x) = \frac{1}{2}\sigma^2(x,\mu)\frac{d^2}{dx^2}f(x) + b(x,\mu)\frac{d}{dx}f(x). \tag{4.2}$$

Here, $\sigma : \mathbb{R} \times \mathcal{P}(\mathbb{R}^{(m+1)}) \mapsto \mathbb{R}$, and $b : \mathbb{R} \times \mathcal{P}(\mathbb{R}^{(m+1)}) \mapsto \mathbb{R}$, are functions such that for each $\mu \in \mathcal{P}(\mathbb{R}^{(m+1)})$ the mappings:

$$x \to \sigma(x,\mu),$$
$$x \to b(x,\mu),$$

are globally Lipchitz continuous and bounded.

Roughly speaking, our **nonlinear martingale problem** consists of finding a stochastic process X such that

$$f(X_t) - f(X_s) - \int_s^t A[\mu(X)]f(X_u)du, \qquad t \geqslant s,$$

is a $\mathcal{F}^X$-martingale for each f in the common domain $\mathcal{D}$. The notation $\mu(X)$ means that the measure depends on the process X. In particular, if $\mu(X) = \mathcal{L}(X_u)$, then we are in the case of McKean nonlinearity. In this work, we are going to consider $\mu(X)$ to be finite-dimensional distributions of the process X. More precisely, we are going to consider

$$\mu(X) = \mathcal{L}(X_{u-\pi_m}, X_{u-\pi_{m-1}}, \cdots, X_{u-\pi_1}, X_u),$$

where $0 \leqslant \pi_1 < \pi_2 < \cdots < \pi_{m-1} < \pi_m$.

4.2 Setting the Nonlinear Martingale Problem

To set our nonlinear martingale problem rigorously we need to specify three components: (1) a collection of points whose purpose is twofold: determining the kind of finite-dimensional distributions which are going to be considered in the nonlinearity, and defining the period of time on which we require an initial condition; (2) a family of operators which describes the evolution of the marginal probability distributions of the process after the initial period of time; and (3) an initial condition which characterises the process on the initial period of time.

Definition 4.1. Let $(\Pi, A[\mu], X^0)$ be a **triplet** where

(1) Π is a collection of points of the form

$$\Pi := \{0 = \pi_0 < \pi_1 < \cdots < \pi_{m-1} < \pi_m = \pi\}.$$

Hereafter, Π will be referred to as a collection of delay points.

(2) $A[\mu] := \{A[\mu] : \mu \in \mathcal{P}(\mathbb{R}^{(m+1)})\}$ is a family of operators of the form

$$A[\mu]f(x) = \frac{1}{2}\sigma^2(x,\mu)\frac{d^2}{dx^2}f(x) + b(x,\mu)\frac{d}{dx}f(x).$$

(3) $X^0 = \{X_t^0 : 0 \leqslant t \leqslant \pi\}$ is an $\mathbb{R}$-valued stochastic process which will play the role of initial condition. Notice that instead of being a single random variable, our problem requires to define the whole process on the initial interval $[0, \pi]$ to be completely defined.

Definition 4.2. A stochastic process $X = \{X_t, t \geqslant 0\}$, defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$, is said to be **a solution of the nonlinear martingale problem for the triplet** $(\mathbf{\Pi}, \mathbf{A}[\mu], \mathbf{X}^0)$ if the process

$$M_t^f = f(X_t) - f(X_\pi) - \int_\pi^t A[\mathcal{L}(X_{\Pi_s})]f(X_s)ds, \qquad t \geqslant \pi, \tag{4.3}$$

where $\mathcal{L}(X_{\Pi_s})$ denotes the marginal distribution of the delay random vector X_Π (as it was defined in Section 3.2.2), is a martingale with respect to the filtration $\mathcal{F}^X$ for all $f \in C_b^2(\mathbb{R})$, and $\mathbb{P} \circ X_t^{-1} = \mathbb{P} \circ X_t^{0-1}$, for every $t \in [0, \pi]$.

Our main result provide sufficient conditions to guarantee the existence and uniqueness of a solution to this type of non-linear martingale problems.

4.3 Solution to the Martingale Problem

The purpose of this section is to show how to construct a solution to the martingale problem for a given triplet $(\Pi, A[\mu], X^0)$. In fact, we are going to prove that the following conditions are sufficient to guarantee existence of a unique solution.

Assumption 4.1. (Global Lipschitz Condition) Both coefficients are uniformly (with respect to μ) Lipschitz continuous with respect to x. That is, there exists a constant $K > 0$ such that

$$|b(x, \mu) - b(y, \mu)| + |\sigma(x, \mu), \sigma(y, \mu)| \leqslant K|x - y|,$$

for every $\mu \in \mathcal{P}(\mathbb{R}^{m+1})$, $x \in \mathbb{R}$, $y \in \mathbb{R}$.

Assumption 4.2. (Linear Growth Condition) There exists a constant $C > 0$ such that

$$|b(x, \mu)|^2 + |\sigma(x, \mu)|^2 \leqslant C^2(1 + |x|^2),$$

for every $\mu \in \mathcal{P}(\mathbb{R}^{m+1})$, $x \in \mathbb{R}$.

Assumption 4.3. (Bounded Condition) There exists a constant $M > 0$ such that

$$\sup_{x \in \mathbb{R}} \{|b(x, \mu)| + |\sigma(x, \mu)|\} \leqslant M, \tag{4.4}$$

for every $\mu \in \mathcal{P}(\mathbb{R}^{m+1})$.

Assumption 4.4. (Initial Process Condition) The process X^0 is continuous and satisfies the following condition

$$\sup_{0 \leqslant t \leqslant \pi} \mathbb{E}[|X_t^0|^4] < \infty.$$

In addition, we require the following extra assumptions:

Assumption 4.5. (Diffusion Coefficient Lipschitz Continuity in Both Variables) $\sigma^2(x, \mu)$ is Lipschitz continuous. That is, there exists a constant $\lambda_\sigma > 0$ such that

$$|\sigma^2(x, \mu) - \sigma^2(\tilde{x}, \tilde{\mu})| \leqslant \lambda_\sigma(|x - \tilde{x}| + |\mu - \tilde{\mu}|_{D^*}), \tag{4.5}$$

for all $x, \tilde{x} \in \mathbb{R}$, and $\mu, \tilde{\mu} \in \mathcal{P}(\mathbb{R}^{m+1})$, where

$$|\mu|_{D^*} = \sup_{|f|_D \leqslant 1} |\langle f, \mu \rangle| = \sup_{|f|_D \leqslant 1} \left| \int f(x) \mu(dx) \right|,$$

and $D = C_\infty^2(\mathbb{R}^{m+1})$.

Assumption 4.6. (Drift Coefficient Lipschitz Continuity in Both Variables) The function $b(x, \mu)$ is Lipschitz continuous. That is, there exists a constant $\lambda_b > 0$ such that

$$|b(x, \mu) - b(\tilde{x}, \tilde{\mu})| \leqslant \lambda_b(|x - \tilde{x}| + |\mu - \tilde{\mu}|_{D^*}), \tag{4.6}$$

for all $x, \tilde{x} \in \mathbb{R}$, and $\mu, \tilde{\mu} \in \mathcal{P}(\mathbb{R}^{m+1})$.

Now, we are ready to present our main existence result.

Theorem 4.1. *Let* $(\Pi, A[\mu], X^0)$ *be a triplet as before. If assumptions 4.1 to 4.6 hold, then there exists a unique solution to the corresponding nonlinear martingale problem.*

Before explaining the main ideas of the proof, let us state some classic results that will help us to construct the solution. The following result can be found as Theorem 2.9 in Karatzas and Shreve, 1991 with slightly different notation.

Lemma 4.2. *Consider the following SDE*

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t, \tag{4.7}$$

Suppose that the coefficients $b(x,t)$*, and* $\sigma(x,t)$ *satisfy the global Lipschitz and linear growth conditions. On some probability space* $(\Omega, \mathcal{F}, \mathbb{P})$*, let* X_0 *be and* $\mathbb{R}$*-valued random variable, independent of the Brownian motion* W*, and with finite second moment:*

$$\mathbb{E}[|X_0|^2] < \infty.$$

Then there exists a continuous, adapted process X *which is a strong solution of equation (4.7) with initial condition* X_0*. Moreover, this process is square-integrable: for every* $T > 0$*, there exists a constant* C *depending only on* K *and* T *such that*

$$\mathbb{E}[|X_t|^2] \leqslant C(1 + \mathbb{E}[|X_0|^2])e^{Ct}; \qquad 0 \leqslant t \leqslant T.$$

Proof. See Theorem 2.9 in Karatzas and Shreve (1991). $\square$

4.3.1 Outline of the Proof

In order to make our proof clearer we decided to include an outline as follows.

Step 1. (Construction of the Approximating Sequence) First, we introduce a sequence of stochastic processes $(X^{(n)})_{n\geqslant 0}$ as follows.

For $n=0$, simply consider the process $X^{(0)}=\{X_t^{(0)}, t\geqslant 0\}$ given by

$$X_t^{(0)} := \begin{cases} X_t^0 & \text{if } 0\leqslant t\leqslant \pi, \\ X_\pi^0 & \text{if } t\geqslant \pi. \end{cases}$$

Now, we are going to use this process to define the next element of the sequence.

For $n=1$, let us introduce the family of probability measures $\nu_\cdot^0$ given by

$$\nu_t^0 = \mathcal{L}(X_{\Pi_t}^{(0)}) = \mathcal{L}(X_{t-\pi}^{(0)}, X_{t-\pi_{m-1}}^{(0)}, \cdots, X_{t-\pi_1}^{(0)}, X_t^{(0)}), \qquad t\geqslant \pi.$$

Then set up the stochastic differential equation

$$dX_t = \sigma(X_t, \nu_t^0)dW_t + b(X_t, \nu_t^0)dt, \qquad t\geqslant \pi,$$

with initial condition $X_\pi = X_\pi^0$. Under assumptions 4.1, 4.2 and 4.4, Lemma 4.2 implies that there exists a continuous, adapted, and square integrable process X which is a unique strong solution of this equation. Besides, it has been proven (see e.g. Stroock and Varadhan, 1972) that if the solution to the SDE exists, then the solution to its corresponding martingale problem also exists; and that if the solution to the SDE is unique, then so is the solution to the martingale problem. Moreover, both solutions are the same in the sense that the solution to the martingale problem is the distribution of the solution to the SDE. So, we are going to call such process $X^{(1)}$.

For $n=2$, we define the family of measures $\nu_\cdot^1$ given by

$$\nu_t^1 = \mathcal{L}(X_{\Pi_t}^{(1)}), \qquad t\geqslant \pi;$$

and set up the SDE

$$dX_t = \sigma(X_t, \nu_t^1)dW_t + b(X_t, \nu_t^1)dt, \qquad t \geqslant \pi,$$

with initial condition $X_\pi = X_\pi^1$. Then by using the same argument as in the construction of the process $X^{(1)}$, we can define $X^{(2)}$ as the unique strong solution the corresponding SDE and whose distribution is the unique solution to the corresponding martingale problem.

Repeating this process, we have defined our sequence of processes $(X^{(n)})_{n\geqslant 0}$. In summary, for $n \geqslant 1$, the process $X^{(n)}$ is defined as the unique strong solution to the SDE give by

$$dX_t^{(n)} = \sigma(X_t^{(n)}, \nu_t^{n-1})dW_t + b(X_t^{(n)}, \nu_t^{n-1})dt, \qquad t \geqslant \pi,$$

where $\nu_\cdot^{n-1}$ is the family of probability measures given by

$$\nu_t^{n-1} = \mathcal{L}(X_{\Pi_t}^{(n-1)}), \qquad t \geqslant \pi;$$

and with initial condition $X_t^n = X_t^0$ for all $t \in [0, \pi]$. That is, $X^{(n)}$ is a process with continuous sample paths such that

$$X_t^{(n)} = X_\pi^0 + \int_\pi^t b(X_s^{(n)}, \nu_s^{n-1})ds + \int_\pi^t \sigma(X_s^{(n)}, \nu_s^{n-1})dW_s, \qquad a.s., \tag{4.8}$$

and $\mathbb{P}(X_t^{(n)} = X_t^0) = 1$ for all $t \in [0, \pi]$.

Besides, the distribution of $X^{(n)}$ is the unique solution to the corresponding martingale problem. That is, for each $f \in C_b^2(\mathbb{R})$ the process

$$f(X_t^{(n)}) - f(X_\pi^{(n)}) - \int_\pi^t A[\nu_s^{n-1}]f(X_s^{(n)})ds, \qquad t \geqslant \pi, \tag{4.9}$$

is a martingale with respect to the filtration generated by $X^{(n)}$; and $X_\pi^{(n)} = X_\pi^0$ almost surely.

It is important to mention that at repeating this process, we have simultaneously built a sequence of curves of distributions, namely

$$\nu_\cdot^n = \{\mathcal{L}(X^{(n-1)}_{\Pi_t}) : t \geqslant \pi\} \qquad n \geqslant 1. \tag{4.10}$$

Our entire construction is illustrated in the following diagram.

$$\begin{tikzcd}
X^0 \arrow[r] & X^{(0)} \arrow[dl] \\
\nu_\cdot^0 = \mathcal{L}(X^{(0)}_{\Pi_\cdot}) \arrow[r] & X^{(1)} \arrow[dl] \\
\nu_\cdot^1 = \mathcal{L}(X^{(1)}_{\Pi_\cdot}) \arrow[r] & X^{(2)} \arrow[dl] \\
\nu_\cdot^2 = \mathcal{L}(X^{(2)}_{\Pi_\cdot}) \arrow[r] & X^{(3)} \\
\cdots & \cdots \\
\nu_\cdot^{n-2} = \mathcal{L}(X^{(n-2)}_{\Pi_\cdot}) \arrow[r] & X^{(n-1)} \arrow[dl] \\
\nu_\cdot^{n-1} = \mathcal{L}(X^{(n-1)}_{\Pi_\cdot}) \arrow[r] & X^{(n)} \arrow[dl] \\
\nu_\cdot^n = \mathcal{L}(X^{(n)}_{\Pi_\cdot}) \arrow[r] & X^{(n+1)}
\end{tikzcd}$$

Fig. 4.1: Approximation sequence from Step 1. Observe that each level of the sequence is characterised by two stochastic processes, namely the current process $X^{(n)}$ and the previous one $X^{(n-1)}$ which is needed to define ν^{n-1}.

Step 2. (Tightness of the Sequence of Processes) We are going to prove that the sequence of processes $(X^{(n)})_{n\geqslant 0}$ is tight. As a consequence we obtain the existence of a subsequence $(X^{(n_k)})_{k\geqslant 0}$ which converges weakly to a limit process X^∞.

Step 3. (Weak convergence of the Sequence of Curves of Distributions) We are going to demonstrate that the sequence of curves of probability measures given in equation (4.10) converges, with respect to a certain metric in a space of curves of probability measures. For this, a contraction argument and the Banach Fixed point Theorem will be used. This convergence, will guarantee the uniqueness of the solution to the nonlinear martingale problem but also will be used to prove the last step of the proof.

Step 4. (Limit Process Solves the Martingale Problem) To conclude, we are going to show that there exists a limit process X^∞, which solves the nonlinear martingale problem corresponding to the triplet $(\Pi, A[\mu], X^0)$. To do this, we will use the fact that tightness of the sequence $(X^{(n)})_{n\geqslant 0}$, obtained in **Step 2**, implies the tightness of the sequence of pairs $(X^{(n)}, X^{(n-1)})_{n\geqslant 0}$. Then, using Prohorov's Theorem and the so-called Skorohod representation theorem we obtain a weakly convergent subsequence $(X^{(n_k)}, X^{(n_k-1)})_{k\geqslant 0}$ with limit process $(X^\infty, \tilde{X}^\infty)$. Finally, we will show that X^∞ solves the martingale problem using the fact that the processes X^∞, $\tilde{X}^\infty$ are "the same" in a certain sense thanks to the convergence result proved in **Step 3**.

■

Now that the general strategy of the proof has been described, we fill in the details.

4.3.2 Existence and Uniqueness

As we mentioned in the outline, assumptions 4.1, 4.2 and 4.4 are sufficient to guarantee the existence and uniqueness of each stochastic process $X^{(n)}$ by Lemma 4.2. So, the sequence of continuous stochastic processes $(X^{(n)})_{n\geqslant 1}$ is well-defined and **Step 1** is complete.

Next, we are going to prove **Step 2**. That is, that the sequence of processes $(X^{(n)})_{n\geqslant 0}$ is tight. To do this, we are going to use the following criterion due to Aldous.

Theorem 4.3. *(Aldous Criterion for Tightness) Let $(X^{(n)})_{n\geqslant 0}$ be a sequence of stochastic processes with càdlàg (French: "continue à droite, limite à gauche") paths. Then the probability measures induced, on the Skorohod space*[1]*, by $(X^{(n)})_{n\geqslant 0}$ are tight if the following two conditions hold for each $T > 0$:*

(i) Stochastically Boundedness. That is

$$\lim_{a\to\infty} \sup_{n} \mathbb{P}(|X^{(n)}|_T \geqslant a) = 0.$$

(ii) For each $\epsilon > 0, \eta > 0$, $m > 0$ there exist constants δ_0 and n_0 such that if $\delta \leqslant \delta_0$ and $n \geqslant n_0$, and if τ is a discrete stopping with respect to $\mathcal{F}^{X^{(n)}}$ satisfying $\tau \leqslant m$, then

$$\mathbb{P}(|X^{(n)}_{\tau+\delta} - X^{(n)}_{\tau}| \geqslant \epsilon) \leqslant \eta.$$

Proof. See Theorem 16.10 in Billingsley (1999). □

[1] The collection of càdlàg functions $f : [0,\infty) \to \mathbb{R}$ is known as Skorohod space after the Soviet mathematician Anatoliy Skorohod.

Thus, we are going to start by proving that the sequence of processes defined above is stochastically bounded. This will follow of course from the boundedness of the coefficients and the fact that the initial condition has bounded moments.

Lemma 4.4. *The sequence of stochastic processes* $(X^{(n)})_{n\geqslant 0}$ *is stochastically bounded. That is,*

$$\lim_{a\to\infty}\sup_{n}\mathbb{P}(|X^{(n)}|_T \geqslant a) = 0,$$

holds for each $T \geqslant 0$.

Proof. For $T \geqslant 0$ and $n \geqslant 0$, we have

$$\mathbb{P}(|X^{(n)}|_T \geqslant a) = \mathbb{P}\left(\sup_{t\in[\pi,T]}|X_t^{(n)}| \geqslant a\right), \tag{4.11}$$

by definition. In order to study the right hand side of this equation, let us introduce the sequence of functions $(\phi_m)_{m=1}^{\infty}$ belonging to $C^{\infty}(\mathbb{R})$ such that

$$\begin{cases} \phi_m(x) = |x| & \text{whenever } \ |x| \geqslant \frac{1}{m}, \\ |x| \leqslant \phi_m(x) \leqslant \dfrac{1}{m}, & \text{whenever } \ |x| \leqslant \frac{1}{m}. \end{cases}$$

Observe that $\phi_m(x)$ increasingly converge to $\uparrow |x|$ as m goes to infinity. Besides, for all $m \geqslant 1$ we have

$$\left|\phi_m^{'}(x)\right| = 1, \quad \text{and} \quad \left|\phi_m^{''}(x)\right| = 0, \quad \forall |x| \geqslant \frac{1}{m},$$

and therefore

$$|\phi_m^{'}(x)| \to 1, \quad \text{and} \quad |\phi_m^{''}(x)| \to 0,$$

as m goes to infinity.

On the other hand, for each $n \geqslant 1$ and $m \geqslant 1$, Ito's formula implies that

$$\phi_m(X_t^{(n)}) = \phi_m(X_\pi^{(n)}) + \int_\pi^t A[\nu_s^{n-1}]\phi_m(X_s^{(n)})ds + \int_\pi^t \phi_m^{'}(X_s^{(n)})\sigma(X_s^{(n)}, \nu_s^{n-1})dW_s,$$

for all $t \geqslant \pi$. In fact, the integral

$$M_t^{(m,n)} = \int_\pi^t \phi_m'(X_s^{(n)})\sigma(X_s^{(n)}, \nu_s^{n-1})dW_s,$$

which can be written as

$$M_t^{(m,n)} = \phi_m(X_t^{(n)}) - \phi_m(X_\pi^{(n)}) - \int_\pi^t A[\nu_s^{n-1}]\phi_m(X_s^{(n)})ds, \qquad t \geqslant \pi,$$

is a local martingale. That is, there exists an increasing sequence of stopping times $(\tau_k \geqslant \pi)_{k\geqslant 0}$ such that the following conditions hold:

- the sequence diverges almost surely, i.e. $\mathbb{P}(\tau_k \to \infty) = 1$;
- for each $k \geqslant 0$, the stopped process

$$M_{t\wedge\tau_k}^{(m,n)} = \phi_m(X_{t\wedge\tau_k}^{(n)}) - \phi_m(X_\pi^{(n)}) - \int_\pi^{t\wedge\tau_k} A[\nu_s^{n-1}]\phi_m(X_s^{(n)})ds, \qquad t \geqslant \pi,$$

  is a martingale.

Hence, for each $n \geqslant 1, m \geqslant 1$ and $k \geqslant 1$, Doob's maximal inequality implies that

$$\mathbb{P}\left(\sup_{t\in[\pi,T]} |M_{t\wedge\tau_k}^{(m,n)}| \geqslant a\right) \leqslant \frac{\mathbb{E}\left[|M_{T\wedge\tau_k}^{(m,n)}|\right]}{a},$$

for any constant $a > 0$. Consequently

$$\begin{aligned}\mathbb{P}\left(\sup_{t\in[\pi,T]} |\phi_m(X_{t\wedge\tau_k}^{(n)})| \geqslant a\right) \leqslant &\frac{1}{a}\mathbb{E}[|M_{T\wedge\tau_k}^{(m,n)}|] + \frac{1}{a}\mathbb{E}[|\phi_m(X_\pi^{(n)})|] \\ &+ \frac{1}{a}\mathbb{E}\left[\int_\pi^{T\wedge\tau_k} |A[\nu_s^{(n-1)}]\phi_m(X_s^{(n)})|ds\right].\end{aligned}$$

Moreover, noticing that $\phi_m(x) \geqslant |x|$ for all x and for any $m \geqslant 1$ by construction, we obtain

$$\mathbb{P}\left(\sup_{t\in[\pi,T]} |X^{(n)}_{t\wedge\tau_k}| \geqslant a\right) \leqslant \frac{1}{a}\mathbb{E}[|M^{(m,n)}_{T\wedge\tau_k}|] + \frac{1}{a}\mathbb{E}[\phi_m(X^{(n)}_{\pi})] \qquad (4.12)$$
$$+ \frac{1}{a}\mathbb{E}\left[\int_{\pi}^{T\wedge\tau_k} \left|A[\nu^{(n-1)}_s]\phi_m(X^{(n)}_s)\right| ds\right].$$

So, let us work on the three terms at the right hand side of (4.12). For the first term, notice that Burkholder-Davis-Gundy inequality implies

$$\mathbb{E}[|M^{(m,n)}_{T\wedge\tau_k}|] \leqslant C_1\mathbb{E}\left[\langle M^{(m,n)}\rangle_{T\wedge\tau_k}\right],$$

for some constant $C_1 > 0$. Next, observe that the martingale $M^{(m,n)}_t$ satisfies

$$\mathbb{E}[\langle M^{(m,n)}\rangle_t] = \mathbb{E}\left[\int_{\pi}^{t} (\phi'_m(X^{(n)}_s))^2\sigma^2(X^{(n)}_s, \nu^{(n-1)}_s)ds\right],$$

and then

$$\frac{1}{a}\mathbb{E}[|M^{(m,n)}_{T\wedge\tau_k}|] \leqslant \frac{C_1}{a}\mathbb{E}\left[\int_{\pi}^{T\wedge\tau_k} (\phi'_m(X^{(n)}_s))^2\sigma^2(X^{(n)}_s, \nu^{(n-1)}_s)ds\right] \qquad (4.13)$$
$$\leqslant \frac{C_1}{a}M^2\mathbb{E}\left[\int_{\pi}^{T\wedge\tau_k} (\phi'_m(X^{(n)}_s))^2 ds\right],$$

where the last inequality follows by using the fact that σ is bounded by assumption 4.3, and the way that we defined the functions ϕ_m. For the second term, we have

$$\frac{1}{a}\mathbb{E}[\phi_m(X^{(n)}_{\pi})] = \frac{1}{a}\mathbb{E}[\phi_m(X^{(0)}_{\pi})] \leqslant \frac{1}{a}\left(1 + \mathbb{E}[|X^{(0)}_{\pi}|]\right), \qquad (4.14)$$

since all the processes have the same distribution at the time π, and using the definition of the functions ϕ_m. Finally, for the last term on the right hand side of (4.12), we have

$$\begin{aligned}
&\frac{1}{a}\mathbb{E}\left[\int_{\pi}^{T\wedge\tau_k}\left|A[\nu_s^{(n-1)}]\phi_m(X_s^{(n)})\right| ds\right] \\
&= \frac{1}{a}\mathbb{E}\left[\int_{\pi}^{T\wedge\tau_k}\left|\frac{1}{2}\sigma^2(X_s^{(n)},\nu_s^{(n-1)})\phi_m''(X_s^{(n)}) + b(X_s^{(n)},\nu_s^{(n-1)})\phi_m'(X_s^{(n)})\right| ds\right],
\end{aligned}$$

and then assumption 4.3 implies that

$$\begin{aligned}
&\frac{1}{a}\mathbb{E}\left[\int_{\pi}^{T\wedge\tau_k}\left|A[\nu_s^{(n-1)}]\phi_m(X_s^{(n)})\right| ds\right] \\
&\leqslant \frac{1}{a}\mathbb{E}\left[\int_{\pi}^{T\wedge\tau_k}\left\{\frac{M^2}{2}\left|\phi_m''(X_s^{(n)})\right| + M\left|\phi_m'(X_s^{(n)})\right|\right\} ds\right].
\end{aligned} \tag{4.15}$$

Together equations (4.12) to (4.15) imply

$$\begin{aligned}
\mathbb{P}\left(\sup_{t\in[\pi,T]}|X_{t\wedge\tau_k}^{(n)}| \geqslant a\right) \leqslant{}& \frac{C_1}{a}M^2\mathbb{E}\left[\int_{\pi}^{T\wedge\tau_k}(\phi_m'(X_s^{(n)}))^2 ds\right] + \frac{1}{a}\left(1+\mathbb{E}[|X_\pi^{(0)}|]\right) \\
&+ \frac{1}{a}\mathbb{E}\left[\int_{\pi}^{T\wedge\tau_k}\left\{\frac{M^2}{2}\left|\phi_m''(X_s^{(n)})\right| + M\left|\phi_m'(X_s^{(n)})\right|\right\} ds\right],
\end{aligned}$$

and taking the limit as both m and k go to infinity, we obtain

$$\begin{aligned}
\mathbb{P}\left(\sup_{t\in[\pi,T]}|X_t^{(n)}| \geqslant a\right) \leqslant{}& \frac{C_1}{a}M^2(T-\pi) + \frac{1}{a}\left(1+\mathbb{E}[|X_\pi^{(0)}|]\right) \\
&+ \frac{M}{a}(T-\pi),
\end{aligned}$$

and the result follows by noticing that the right hand side does not depend on n and taking the limit as a tends to infinity. □

In order to prove that our sequence of processes $(X^{(n)})_{n\geqslant 0}$, as defined in **Step 1**, satisfies the second condition in the Aldous Criterion for Tightness (Theorem 4.3), we are going to use the following equivalent condition.

[A] Aldous'Condition: A sequence $(X^{(n)})_{n\geqslant 0}$, of $\mathbb{R}$-valued stochastic processes, each defined on its own probability space with a fixed filtration $\mathcal{F}^{(n)}$, is said to satisfy the Aldous'condition if

$$\left|X^{(n)}_{\tau_n+h_n} - X^{(n)}_{\tau_n}\right| \to 0 \qquad \text{in probability, as} \quad n \to \infty,$$

for any sequence of bounded $\mathcal{F}^{(n)}$-stopping times $(\tau_n)_{n\geqslant 0}$ and any sequence of positive numbers $(h_n)_{n\geqslant 0}$ such that $h_n \to 0$, as $n \to \infty$.

The proof of this condition being equivalent to the condition (ii) in Theorem 4.3 can be found, for instance, in Proposition 4.8.1 in Kolokoltsov (2011). So, the next result shows that our sequence satisfies the Aldous Condition.

Lemma 4.5. *The sequence of continuous stochastic processes* $(X^{(n)})_{n\geqslant 0}$ *constructed above (see **Step 1** and Figure 4.1.) satisfies the Aldous Condition [A].*

Proof. Let $(\tau_n)_{n\geqslant 0}$ be a sequence of bounded $\mathcal{F}^{(n)}$-stopping times; without loss of generality we are going to assume that

$$\pi \leqslant \tau_n \leqslant T, \qquad \text{a.s.}$$

Besides, let $(h_n)_{n\geqslant 0}$ a sequence of positive numbers such that $h_n \to 0$, as $n \to \infty$. Let $\epsilon > 0$, we want to prove that for all $\epsilon > 0$

$$\lim_{n\to\infty} P(|X^{(n)}_{\tau_n+h_n} - X^{(n)}_{\tau_n}| \geqslant \epsilon) = 0. \tag{4.16}$$

First of all remind that for each $n \geqslant 1$,

$$X^{(n)}_t = X^{(0)}_\pi + \int_\pi^t b(X^{(n)}_s, \nu^{n-1}_s)ds + \int_\pi^t \sigma(X^{(n)}_s, \nu^{n-1}_s)dW_s, \qquad t \geqslant \pi,$$

by definition. Then we have

$$
\begin{aligned}
&\mathbb{E}[|X^{(n)}_{\tau_n+t} - X^{(n)}_{\tau_n}|^2] \\
&= \mathbb{E}\left[\left|\int_{\tau_n}^{\tau_n+t} b(X^{(n)}_s, \nu^{n-1}_s)ds + \int_{\tau_n}^{\tau_n+t} \sigma(X^{(n)}_s, \nu^{n-1}_s)dW_s\right|^2\right] \\
&\leqslant 2\mathbb{E}\left[\left(\int_{\pi}^{t} b(X^{(n)}_{\tau_n+s}, \nu^{n-1}_{\tau_n+s})ds\right)^2\right] + 2\mathbb{E}\left[\left(\int_{\pi}^{t} \sigma(X^{(n)}_{\tau_n+s}, \nu^{n-1}_{\tau_n+s})dW_s\right)^2\right] \\
&\leqslant 2(t-\pi)\mathbb{E}\left[\int_{\pi}^{t} b^2(X^{(n)}_{\tau_n+s}, \nu^{n-1}_{\tau_n+s})ds\right] + 2\mathbb{E}\left[\int_{\pi}^{t} \sigma^2(X^{(n)}_{\tau_n+s}, \nu^{n-1}_{\tau_n+s})ds\right] \\
&\leqslant 2(t-\pi)^2M^2 + 2M^2(t-\pi) \\
&= 2(t-\pi)M^2(1+(t-\pi)),
\end{aligned}
$$

for all $t \geqslant 0$, by 4.3 and Holder inequality.

Hence, by Markov inequality it follows that

$$
P(|X^{(n)}_{\tau_n+h_n} - X^{(n)}_{\tau_n}| \geqslant \epsilon) \leqslant \frac{1}{\epsilon^2}2M^2(h_n-\pi)(1+h_n-\pi),
$$

and the results follows by taking the limit as n tends to infinity. □

In summary, Theorem 4.3 together with Lemmas 4.4 and 4.5 imply that the sequence of stochastic processes $(X^{(n)})_{n\geqslant 0}$, as defined in **Step 1**, is tight. Thus, Prohorov's Theorem implies that it contains a weakly convergent subsequence $(X^{(n_k)})_{k\geqslant 0}$. Moreover, by the so-called Skorohod representation theorem (see e.g. Theorem 3.1.8 in Ethier and Kurtz, 1986), we may assume that the processes $(X^{(n_k)})_{k\geqslant 0}$ and the limiting process X^∞ are defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$ where

$$
X^{(n_k)} \to X^\infty \quad \text{almost surely as} \quad k \to \infty. \tag{4.17}
$$

That is,

$$
\lim_{k\to\infty} \mathbb{E}[f(X^{(n_k)})] = \mathbb{E}[f(X^\infty)],
$$

for all f bounded and continuous function defined on the Skorohod space.

Remark 4.1. Remember that convergence in the Skorohod space implies convergence of finite dimensional distributions for all continuity points (see e.g.Theorem 3.7.8 in Ethier and Kurtz, 1986) of the limiting process. Besides, the limiting process is stochastically (or in probability) continuous as a consequence of satisfying Aldous condition (Proposition 4.8.2 in Kolokoltsov, 2011). Hence, for every finite set $\{t_0, t_1 \cdots, t_N\} \subset [0, \infty]$ we have the following weak convergence

$$\mathcal{L}(X_{t_0}^{(n_k)}, X_{t_1}^{(n_k)}, \cdots, X_{t_N}^{(n_k)}) \to \mathcal{L}(X_{t_0}^{\infty}, X_{t_1}^{\infty}, \cdots, X_{t_N}^{\infty}) \quad \text{as} \quad k \to \infty.$$

That is, for every finite set $\{t_0, t_1 \cdots, t_N\} \subset [0, \infty]$ we have

$$\lim_{k\to\infty} \mathbb{E}[f(X_{t_0}^{(n_k)}, X_{t_1}^{(n_k)}, \cdots, X_{t_N}^{(n_k)})] = \mathbb{E}[f(X_{t_0}^{\infty}, X_{t_1}^{\infty}, \cdots, X_{t_N}^{\infty})], \qquad a.s., \tag{4.18}$$

for every f bounded.

Notice that in particular, we have

$$\lim_{k\to\infty} \mathbb{E}[f(X_t^{(n_k)})] = \mathbb{E}[f(X_t^{\infty})], \qquad a.s., \tag{4.19}$$

for any $f \in C_b(\mathbb{R})$ and $t \geqslant 0$.

Now, we are proceed to complete **Step 3**. In order to study the convergence of the sequence defined in (4.10) we need to specify the metric that will be used. To do this, notice that each element of the sequence $\nu^n_{\cdot}$ has the same initial distribution by construction. So, let us set $\xi = \mathcal{L}(X_{\Pi_\pi}^{(0)})$, and consider the space of curves of probability $C_\xi([\pi, T], \mathcal{P}(\mathbb{R}^{m+1}))$ as defined at the end of Section 2.1. Then, fix $T > \pi$ and let us introduce the metric

$$\|\nu_\cdot\|_T = \sup_{\pi \leqslant t \leqslant T} |\nu_t|,$$

where

$$|\nu_t| = \sup_{|f|_D \leqslant 1} \left| \int f(x)\nu_t(dx) \right|,$$

and D is the space of continuously twice differentiable functions vanishing at infinity. So, by definition we have

$$\|\nu_\cdot - \mu_\cdot\|_T = \sup_{\pi \leqslant t \leqslant T} \sup_{|f|_D \leqslant 1} \left| \int f(x)\nu_t(dx) - \int f(x)\mu_t(dx) \right|,$$

for any two curves of probability measures in $C_\xi([\pi, T], \mathcal{P}(\mathbb{R}^{m+1}))$.

So, let us introduce the mapping Φ as the map which associates to a given curve $\nu_\cdot^n$ belonging to the space $C_\xi([\pi, T], \mathcal{P}(\mathbb{R}^{m+1}))$, the curve of probability measures

$$\nu_\cdot^{n+1} = \{\mathcal{L}(X_{\Pi_t}^{(n+1)}) : \pi \leqslant t \leqslant T\}, \tag{4.20}$$

where of course

$$X_t^{(n+1)} = X_\pi^0 + \int_\pi^t b(X_s^{(n+1)}, \nu_s^n)ds + \int_\pi^t \sigma(X_s^{(n+1)}, \nu_s^n)dW_s, \qquad a.s. \tag{4.21}$$

for $n \geqslant 1$, by the way in which the sequence was constructed. Next, observe that if a process $X = \{X_t : \pi \leqslant t \leqslant T\}$ is a solution of the nonlinear SDE

$$dX_t = \sigma(X_t, \mathcal{L}(X_{\Pi_t}))dW_t + b(X_t, \mathcal{L}(X_{\Pi_t}))dt, \qquad \pi \leqslant t \leqslant T,$$

with initial condition $X_t = X_t^0, \ \ \forall t \in [0, \pi]$, then the curve of probability measures $\mathcal{L}(X_{\Pi_\cdot})$ is a fixed point of Φ. Conversely if Φ has a fixed point, i.e., there exists a curve of probability measures ν^∞, such that $\Phi(\nu^\infty) = \nu^\infty$ this will define a process which solves the nonlinear SDE up to time T. So, proving that the mapping Φ has a unique fixed point would guarantee the uniqueness of the solution to the nonlinear martingale problem.

In order to prove that Φ has a unique fixed point, we are going to prove that Φ is a contraction. That is, that there exist a constant $q \in (0,1)$ such that

$$\|\Phi(\nu_\cdot^n) - \Phi(\nu_\cdot^{n-1})\|_T \leqslant q\|\nu_\cdot^n - \nu_\cdot^{n-1}\|_T, \tag{4.22}$$

for any $n \geqslant 1$.

By definition, we have

$$\|\Phi(\nu_\cdot^n) - \Phi(\nu_\cdot^{n-1})\|_T = \sup_{\pi \leqslant t \leqslant T} \sup_{|f|_D \leqslant 1} \left| \mathbb{E}[f(X_{\Pi_t}^{(n+1)})] - \mathbb{E}[f(X_{\Pi_t}^{(n)})] \right|. \tag{4.23}$$

Now, let fixed f in D. Since f is twice differentiable, and assumptions 4.1, 4.3 and 4.4 hold, Theorem 3.7 implies that

$$\begin{aligned}
f({X_{\Pi_t}}^{(n+1)}) =& f({X_{\Pi_\pi}}^{(n+1)}) + \sum_{j=1}^{m+1} \int_{\pi-\pi_{(m-j+1)}}^{t-\pi_{(m-j+1)}} \partial_j f(X^{(n+1)}_{\Pi_{s+\pi_{(m-j+1)}}}) \sigma(X_s^{(n+1)}, \nu_s^n) \delta W_s \\
&+ \sum_{j=1}^{m+1} \int_\pi^t \partial_j f(X_{\Pi_s}^{(n+1)}) b(X^{(n+1)}_{s-\pi_{m-j+1}}, \nu^n_{s-\pi_{m-j+1}}) ds \\
&+ \frac{1}{2} \sum_{j=1}^{m+1} \int_\pi^t \partial^2_{jj} f(X_{\Pi_s}^{(n+1)}) \sigma^2(X^{(n+1)}_{s-\pi_{m-j+1}}, \nu^n_{s-\pi_{m-j+1}}) ds \\
&+ \sum_{k>j} \int_\pi^t \partial^2_{jk} f(X_{\Pi_s}^{(n+1)}) \sigma(X^{(n+1)}_{s-\pi_{m-j+1}}, \nu^n_{s-\pi_{m-j+1}}) D_{s-\pi_{m-j+1}} X^{(n+1)}_{s-\pi_{m-k+1}} ds,
\end{aligned}$$

almost surely.

Analogously, we have

$$\begin{aligned}f(X^{(n)}_{\Pi_t}) =& f(X^{(n)}_{\Pi_\pi}) + \sum_{j=1}^{m+1} \int_{\pi-\pi_{(m-j+1)}}^{t-\pi_{(m-j+1)}} \partial_j f(X^{(n)}_{\Pi_{s+\pi_{(m-j+1)}}})\sigma(X^{(n)}_s, \nu^{n-1}_s)\delta W_s \\
&+ \sum_{j=1}^{m+1} \int_\pi^t \partial_j f(X^{(n)}_{\Pi_s}) b(X^{(n)}_{s-\pi_{m-j+1}}, \nu^{n-1}_{s-\pi_{m-j+1}}) ds \\
&+ \frac{1}{2}\sum_{j=1}^{m+1} \int_\pi^t \partial^2_{jj} f(X^{(n)}_{\Pi_s})\sigma^2(X^{(n)}_{s-\pi_{m-j+1}}, \nu^{n-1}_{s-\pi_{m-j+1}}) ds \\
&+ \sum_{k>j} \int_\pi^t \partial^2_{jk} f(X^{(n)}_{\Pi_s})\sigma(X^{(n)}_{s-\pi_{m-j+1}}, \nu^{n-1}_{s-\pi_{m-j+1}}) D_{s-\pi_{m-j+1}} X^{(n)}_{s-\pi_{m-k+1}} ds,\end{aligned}$$

almost surely.

Besides Lemma 3.6 implies the following boundedness property of the Malliavin derivative

$$\sup_{0\leqslant r\leqslant t} \mathbb{E}\left[\sup_{r\leqslant s\leqslant T} |D_r X^{(n+1)}_s|^p\right] < \infty,$$

and the derivative $D_r X^{(n+1)}_s$ satisfies the following linear SDE

$$D_r X^{(n+1)}_t = \sigma(X^{(n+1)}_r, \nu^n_r) + \int_r^t \tilde{\sigma}_s D_r X^{(n+1)}_s dW_s + \int_r^t \tilde{b}_s D_r X^{(n+1)}_s ds,$$

for all $r \leqslant t$, a.s., and

$$D_r X^{(n+1)}_t = 0,$$

for $r > t$, a.s.; where $\tilde{\sigma}$ and $\tilde{b}$ are uniformly bounded and adapted processes. Analogously, we obtain the same boundedness property for the corresponding Malliavin derivative of the process $X^{(n)}$.

Thus, using the fact that Skorohod integrals have zero expectation, martingale property for the Ito integral, and the boundedness of the Malliavin derivative provided by

Lemma 3.6 if follows that the difference within the absolute value in the right hand side of (4.23) is bounded as follows

$$
\begin{aligned}
\mathbb{E}[f(X_{\Pi_t}^{(n+1)})] - \mathbb{E}[f(X_{\Pi_t}^{(n)})] \leqslant & \int_\pi^t \mathbb{E}\left[\sum_{j=1}^{m+1} \partial_j f(X_{\Pi_s}^{(n+1)}) b(X_{s-\pi_{m-j+1}}^{(n+1)}, \nu_{s-\pi_{m-j+1}}^{n})\right] ds \\
& + \int_\pi^t \mathbb{E}\left[\frac{1}{2}\sum_{j=1}^{m+1} \partial_{jj}^2 f(X_{\Pi_s}^{(n+1)}) \sigma^2(X_{s-\pi_{m-j+1}}^{(n+1)}, \nu_{s-\pi_{m-j+1}}^{n})\right] ds \\
& + \int_\pi^t \mathbb{E}\left[\sum_{j>k} \partial_{jk}^2 f(X_{\Pi_s}^{(n+1)}) \sigma^2(X_{s-\pi_{m-j+1}}^{(n+1)}, \nu_{s-\pi_{m-j+1}}^{n})\right] ds \\
& - \int_\pi^t \mathbb{E}\left[\sum_{j=1}^{m+1} \partial_j f(X_{\Pi_s}^{(n)}) b(X_{s-\pi_{m-j+1}}^{(n)}, \nu_{s-\pi_{m-j+1}}^{n-1})\right] ds \\
& - \int_\pi^t \mathbb{E}\left[\frac{1}{2}\sum_{j=1}^{m+1} \partial_{jj}^2 f(X_{\Pi_s}^{(n)}) \sigma^2(X_{s-\pi_{m-j+1}}^{(n)}, \nu_{s-\pi_{m-j+1}}^{n-1})\right] ds \\
& - \int_\pi^t \mathbb{E}\left[\sum_{j>k} \partial_{jk}^2 f(X_{\Pi_s}^{(n)}) \sigma^2(X_{s-\pi_{m-j+1}}^{(n)}, \nu_{s-\pi_{m-j+1}}^{n-1})\right] ds.
\end{aligned}
$$

Now, let us introduce the following notation in order to be able to simplify the above expression. Consider

$$
\begin{aligned}
L[\nu]f(X_{\Pi_s}^{(n+1)}) = & \sum_{j=1}^{m+1} \partial_j f(X_{\Pi_s}^{(n+1)}) b(X_{s-\pi_{m-j+1}}^{(n+1)}, \nu_{s-\pi_{m-j+1}}) \\
& + \frac{1}{2}\sum_{j=1}^{m+1}\sum_{k=1}^{m+1} \partial_{jk}^2 f(X_{\Pi_s}^{(n+1)}) \sigma^2(X_{s-\pi_{m-j+1}}^{(n+1)}, \nu_{s-\pi_{m-j+1}}),
\end{aligned}
$$

for any given probability measure ν in $\mathcal{P}(\mathbb{R}^{m+1})$. Using this notation, our above expression becomes

$$
\begin{aligned}
& \mathbb{E}[f(X_{\Pi_t}^{(n+1)})] - \mathbb{E}[f(X_{\Pi_t}^{(n)})] \\
& \leqslant \int_\pi^t \mathbb{E}\left[L[\nu^n]f(X_{\Pi_s}^{(n+1)})\right] ds - \int_\pi^t \mathbb{E}\left[L[\nu^{n-1}]f(X_{\Pi_s}^{(n)})\right] ds.
\end{aligned}
$$

Taking absolute value, adding and subtracting one extra term, and using the triangle inequality we obtain

$$|\mathbb{E}[f(X^{(n+1)}_{\Pi_t})] - \mathbb{E}[f(X^{(n)}_{\Pi_t})]| \leqslant I_1(t, f) + I_2(t, f), \tag{4.24}$$

where

$$I_1(t, f) = \int_{\pi}^{t} |\mathbb{E}\left[L[\nu^n]f(X^{(n+1)}_{\Pi_s}) - L[\nu^n]f(X^{(n)}_{\Pi_s})\right]|ds, \tag{4.25}$$

and

$$I_2(t, f) = \int_{\pi}^{t} |\mathbb{E}\left[L[\nu^n]f(X^{(n)}_{\Pi_s}) - L[\nu^{n-1}]f(X^{(n)}_{\Pi_s})\right]|ds. \tag{4.26}$$

Replacing (4.24) in (4.23), we obtain

$$\|\Phi(\nu^n_{\cdot}) - \Phi(\nu^{n-1}_{\cdot})\|_T \leqslant \sup_{\pi \leqslant t \leqslant T} \sup_{|f|_D \leqslant 1} \{I_1(t, f) + I_2(t, f)\}, \tag{4.27}$$

where the integrals I_1 and I_2 are as defined above. Next, we study these integrals.

For the first integral, notice that

$$\begin{aligned}
&L[\nu^n]f(X^{(n+1)}_{\Pi_s}) - L[\nu^n]f(X^{(n)}_{\Pi_s}) \\
&= \sum_{j=1}^{m+1} \partial_j f(X^{(n+1)}_{\Pi_s}) b(X^{(n+1)}_{s-\pi_{m-j+1}}, \nu^n_{s-\pi_{m-j+1}}) \\
&\quad + \frac{1}{2} \sum_{j,k=1}^{m+1} \partial^2_{jk} f(X^{(n+1)}_{\Pi_s}) \sigma^2(X^{(n+1)}_{s-\pi_{m-j+1}}, \nu^n_{s-\pi_{m-j+1}}) \\
&\quad - \sum_{j=1}^{m+1} \partial_j f(X^{(n)}_{\Pi_s}) b(X^{(n)}_{s-\pi_{m-j+1}}, \nu^n_{s-\pi_{m-j+1}}) \\
&\quad - \frac{1}{2} \sum_{j,k=1}^{m+1} \partial^2_{jk} f(X^{(n)}_{\Pi_s}) \sigma^2(X^{(n)}_{s-\pi_{m-j+1}}, \nu^n_{s-\pi_{m-j+1}}),
\end{aligned}$$

which can be written as

$$\begin{aligned}
&L[\nu^n]f(X^{(n+1)}_{\Pi_s}) - L[\nu^n]f(X^{(n)}_{\Pi_s})\\
&= \sum_{j=1}^{m+1}\Big\{\partial_j f(X^{(n+1)}_{\Pi_s})b(X^{(n+1)}_{s-\pi_{m-j+1}},\nu^n_{s-\pi_{m-j+1}}) - \partial_j f(X^{(n)}_{\Pi_s})b(X^{(n)}_{s-\pi_{m-j+1}},\nu^n_{s-\pi_{m-j+1}})\Big\}\\
&\quad + \frac{1}{2}\sum_{j=1}^{m+1}\sum_{k=1}^{m+1}\Big\{\partial^2_{jk}f(X^{(n+1)}_{\Pi_s})\sigma^2(X^{(n+1)}_{s-\pi_{m-j+1}},\nu^n_{s-\pi_{m-j+1}})\\
&\qquad\qquad - \partial^2_{jk}f(X^{(n)}_{\Pi_s})\sigma^2(X^{(n)}_{s-\pi_{m-j+1}},\nu^n_{s-\pi_{m-j+1}})\Big\}.
\end{aligned}$$

Taking the absolute value and integrating from π to t with respect to s, we obtain

$$\begin{aligned}
&\int_\pi^t |\mathbb{E}\left[L[\nu^n]f(X^{(n+1)}_{\Pi_s}) - L[\nu^n]f(X^{(n)}_{\Pi_s})\right]|ds\\
&= \int_\pi^t |\mathbb{E}\Bigg[\sum_{j=1}^{m+1}\Big\{\partial_j f(X^{(n+1)}_{\Pi_s})b(X^{(n+1)}_{s-\pi_{m-j+1}},\nu^n_{s-\pi_{m-j+1}}) - \partial_j f(X^{(n)}_{\Pi_s})b(X^{(n)}_{s-\pi_{m-j+1}},\nu^n_{s-\pi_{m-j+1}})\Big\}\\
&\quad + \frac{1}{2}\sum_{j,k=1}^{m+1}\Big\{\partial^2_{jk}f(X^{(n+1)}_{\Pi_s})\sigma^2(X^{(n+1)}_{s-\pi_{m-j+1}},\nu^n_{s-\pi_{m-j+1}})\\
&\qquad\qquad - \partial^2_{jk}f(X^{(n)}_{\Pi_s})\sigma^2(X^{(n)}_{s-\pi_{m-j+1}},\nu^n_{s-\pi_{m-j+1}})\Big\}\Bigg]|ds\\
&\leqslant \sum_{j=1}^{m+1}\int_\pi^t |\mathbb{E}\left[\Big\{\partial_j f(X^{(n+1)}_{\Pi_s})b(X^{(n+1)}_{s-\pi_{m-j+1}},\nu^n_{s-\pi_{m-j+1}}) - \partial_j f(X^{(n)}_{\Pi_s})b(X^{(n)}_{s-\pi_{m-j+1}},\nu^n_{s-\pi_{m-j+1}})\Big\}\right]|ds\\
&\quad + \frac{1}{2}\sum_{j,k=1}^{m+1}\int_\pi^t |\mathbb{E}\Big[\Big\{\partial^2_{jk}f(X^{(n+1)}_{\Pi_s})\sigma^2(X^{(n+1)}_{s-\pi_{m-j+1}},\nu^n_{s-\pi_{m-j+1}})\\
&\qquad\qquad - \partial^2_{jk}f(X^{(n)}_{\Pi_s})\sigma^2(X^{(n)}_{s-\pi_{m-j+1}},\nu^n_{s-\pi_{m-j+1}})\Big\}\Big]|ds,
\end{aligned}$$

where the last equality was obtained by using the triangle inequality. Now, the key point consist of noticing that for any given probability measure ν, the mappings

$$F_j(x) = \frac{1}{M}\partial_j f(x)b(x_j,\nu), \qquad j = 1\cdots, m+1,$$

and

$$G_{jk}(x) = \frac{1}{M}\partial^2_{jk}f(x)\sigma^2(x_j,\nu), \qquad j,k = 1,\cdots, m+1,$$

where M is the constant from equation (4.4), belongs to D thanks to assumption 4.3. Consequently, we have

$$\begin{aligned}
&\int_\pi^t |\mathbb{E}\left[L[\nu^n]f(X^{(n+1)}_{\Pi_s}) - L[\nu^n]f(X^{(n)}_{\Pi_s})\right]|ds \\
&\leqslant M\sum_{j=1}^{m+1}\int_\pi^t |\langle F_j, \mathcal{L}(X^{(n+1)}_{\Pi_s}) - \mathcal{L}(X^{(n)}_{\Pi_s})\rangle|ds + \frac{1}{2}M\sum_{j,k=1}^{m+1}\int_\pi^t |\langle G_{jk}, \mathcal{L}(X^{(n+1)}_{\Pi_s}) - \mathcal{L}(X^{(n)}_{\Pi_s})\rangle|ds \\
&\leqslant M\sum_{j=1}^{m+1}\int_\pi^t |\mathcal{L}(X^{(n+1)}_{\Pi_s}) - \mathcal{L}(X^{(n)}_{\Pi_s})|ds + \frac{1}{2}M\sum_{j,k=1}^{m+1}\int_\pi^t |\mathcal{L}(X^{(n+1)}_{\Pi_s}) - \mathcal{L}(X^{(n)}_{\Pi_s})|ds \\
&= M\sum_{j=1}^{m+1}\int_\pi^t |\nu^{n+1}_s - \nu^n_s|ds + \frac{1}{2}M\sum_{j,k=1}^{m+1}\int_\pi^t |\nu^{n+1}_s - \nu^n_s|ds \\
&= M(m+1)\int_\pi^t |\nu^{n+1}_s - \nu^n_s|ds + \frac{1}{2}M(m+1)^2\int_\pi^t |\nu^{n+1}_s - \nu^n_s|ds,
\end{aligned}$$

which in turn implies

$$I_1(t,f) \leqslant C_1\int_\pi^t |\nu^{n+1}_s - \nu^n_s|ds,$$

where the constant C_1 only depends on M and m. Then, we can conclude that

$$\sup_{\pi\leqslant t\leqslant T}\sup_{|f|_D\leqslant 1} I_1(t,f) \leqslant C_1(T-\pi)\|\nu^{n+1}_\cdot - \nu^n_\cdot\|_T, \tag{4.28}$$

where the constant C_1 only depends on M and m.

Similarly, to treat the second integral I_2, notice that we have

$$\begin{aligned}
&L[\nu^n]f(X^{(n)}_{\Pi_s}) - L[\nu^{n-1}]f(X^{(n)}_{\Pi_s}) \\
&= \sum_{j=1}^{m+1}\partial_j f(X^{(n)}_{\Pi_s})b(X^{(n)}_{s-\pi_{m-j+1}}, \nu^n_{s-\pi_{m-j+1}}) + \frac{1}{2}\sum_{j,k=1}^{m+1}\partial^2_{jk}f(X^{(n)}_{\Pi_s})\sigma^2(X^{(n)}_{s-\pi_{m-j+1}}, \nu^n_{s-\pi_{m-j+1}}) \\
&\quad - \sum_{j=1}^{m+1}\partial_j f(X^{(n)}_{\Pi_s})b(X^{(n)}_{s-\pi_{m-j+1}}, \nu^{n-1}_{s-\pi_{m-j+1}}) - \frac{1}{2}\sum_{j,k=1}^{m+1}\partial^2_{jk}f(X^{(n)}_{\Pi_s})\sigma^2(X^{(n)}_{s-\pi_{m-j+1}}, \nu^{n-1}_{s-\pi_{m-j+1}}) \\
&= \sum_{j=1}^{m+1}\partial_j f(X^{(n)}_{\Pi_s})\left[b(X^{(n)}_{s-\pi_{m-j+1}}, \nu^n_{s-\pi_{m-j+1}}) - b(X^{(n)}_{s-\pi_{m-j+1}}, \nu^{n-1}_{s-\pi_{m-j+1}})\right] \\
&\quad + \frac{1}{2}\sum_{j,k=1}^{m+1}\partial^2_{jk}f(X^{(n)}_{\Pi_s})\left[\sigma^2(X^{(n)}_{s-\pi_{m-j+1}}, \nu^n_{s-\pi_{m-j+1}}) - \sigma^2(X^{(n)}_{s-\pi_{m-j+1}}, \nu^{n-1}_{s-\pi_{m-j+1}})\right].
\end{aligned}$$

and then

$$
\begin{aligned}
&|L[\nu^n]f(X^{(n)}_{\Pi_s}) - L[\nu^{n-1}]f(X^{(n)}_{\Pi_s})| \\
&\leqslant \sum_{j=1}^{m+1} \left|\partial_j f(X^{(n)}_{\Pi_s}) \left[b(X^{(n)}_{s-\pi_{m-j+1}}, \nu^n_{s-\pi_{m-j+1}}) - b(X^{(n)}_{s-\pi_{m-j+1}}, \nu^{n-1}_{s-\pi_{m-j+1}})\right]\right| \\
&\quad + \frac{1}{2}\sum_{j,k=1}^{m+1} \left|\partial^2_{jk} f(X^{(n)}_{\Pi_s}) \left[\sigma^2(X^{(n)}_{s-\pi_{m-j+1}}, \nu^n_{s-\pi_{m-j+1}}) - \sigma^2(X^{(n)}_{s-\pi_{m-j+1}}, \nu^{n-1}_{s-\pi_{m-j+1}})\right]\right|.
\end{aligned}
$$

Taking the absolute value and integrating from π to t with respec to s, we obtain

$$
\begin{aligned}
&\int_\pi^t |\mathbb{E}\left[L[\nu^n]f(X^{(n)}_{\Pi_s}) - L[\nu^{n-1}]f(X^{(n)}_{\Pi_s})\right]| ds \\
&\leqslant \sum_{j=1}^{m+1} \int_\pi^t \mathbb{E}\left[\left|\partial_j f(X^{(n)}_{\Pi_s}) \left[b(X^{(n)}_{s-\pi_{m-j+1}}, \nu^n_{s-\pi_{m-j+1}}) - b(X^{(n)}_{s-\pi_{m-j+1}}, \nu^{n-1}_{s-\pi_{m-j+1}})\right]\right|\right] ds \\
&\quad + \frac{1}{2}\sum_{j,k=1}^{m+1} \int_\pi^t \mathbb{E}\left[\left|\partial^2_{jk} f(X^{(n)}_{\Pi_s}) \left[\sigma^2(X^{(n)}_{s-\pi_{m-j+1}}, \nu^n_{s-\pi_{m-j+1}}) - \sigma^2(X^{(n)}_{s-\pi_{m-j+1}}, \nu^{n-1}_{s-\pi_{m-j+1}})\right]\right|\right] ds \\
&= \sum_{j=1}^{m+1} \int_{\pi-\pi_{m-j+1}}^{t-\pi_{m-j+1}} \mathbb{E}\left[\left|\partial_j f(X^{(n)}_{\Pi_{s+\pi_{m-j+1}}}) \left[b(X^{(n)}_s, \nu^n_s) - b(X^{(n)}_s, \nu^{n-1}_s)\right]\right|\right] ds \\
&\quad + \frac{1}{2}\sum_{j,k=1}^{m+1} \int_{\pi-\pi_{m-j+1}}^{t-\pi_{m-j+1}} \mathbb{E}\left[\left|\partial^2_{jk} f(X^{(n)}_{\Pi_{s+\pi_{m-j+1}}}) \left[\sigma^2(X^{(n)}_s, \nu^n_s) - \sigma^2(X^{(n)}_s, \nu^{n-1}_s)\right]\right|\right] ds,
\end{aligned}
$$

where the last equality follows by making a change of variable. Then, using the fact that f has bounded partial derivatives, and that assumptions 4.5 and 4.6 hold we obtain

$$
\begin{aligned}
I_2(t,f) &= \int_\pi^t |\mathbb{E}\left[L[\nu^n]f(X^{(n)}_{\Pi_s}) - L[\nu^{n-1}]f(X^{(n)}_{\Pi_s})\right]| ds \\
&\leqslant \sum_{j=1}^{m+1} \int_{\pi-\pi_{m-j+1}}^{t-\pi_{m-j+1}} \mathbb{E}\left[\left|b(X^{(n)}_s, \nu^n_s) - b(X^{(n)}_s, \nu^{n-1}_s)\right|\right] ds \\
&\quad + \frac{1}{2}\sum_{j,k=1}^{m+1} \int_{\pi-\pi_{m-j+1}}^{t-\pi_{m-j+1}} \mathbb{E}\left[\left|\sigma^2(X^{(n)}_s, \nu^n_s) - \sigma^2(X^{(n)}_s, \nu^{n-1}_s)\right|\right] ds \\
&\leqslant \lambda_b \sum_{j=1}^{m+1} \int_{\pi-\pi_{m-j+1}}^{t-\pi_{m-j+1}} \mathbb{E}\left[|\nu^n_s - \nu^{n-1}_s|\right] ds \\
&\quad + \frac{1}{2}\lambda_\sigma \sum_{j,k=1}^{m+1} \int_{\pi-\pi_{m-j+1}}^{t-\pi_{m-j+1}} \mathbb{E}\left[|\nu^n_s - \nu^{n-1}_s|\right] ds,
\end{aligned}
$$

where of course the constants λ_b and λ_σ come from equations (4.5) and (4.6), respectively. So, taking the supremum over t and f we obtain

$$\begin{aligned}\sup_{\pi\leqslant t\leqslant T}\sup_{|f|_D\leqslant 1} I_2(t,f) \leqslant \sup_{\pi\leqslant t\leqslant T}&\left\{\lambda_b\sum_{j=1}^{m+1}\int_{\pi-\pi_{m-j+1}}^{t-\pi_{m-j+1}}\mathbb{E}\left[|\nu_s^n-\nu_s^{n-1}|\right]ds\right.\\&\left.+\frac{1}{2}\lambda_\sigma\sum_{j,k=1}^{m+1}\int_{\pi-\pi_{m-j+1}}^{t-\pi_{m-j+1}}\mathbb{E}\left[|\nu_s^n-\nu_s^{n-1}|\right]ds\right\},\\&\leqslant \lambda_b(m+1)(T-\pi)\|\nu_\cdot^n-\nu_\cdot^{n-1}\|_T+\frac{1}{2}\lambda_\sigma(T-\pi)(m+1)^2\|\nu_\cdot^n-\nu_\cdot^{n-1}\|_T,\end{aligned}$$

which implies that

$$\sup_{\pi\leqslant t\leqslant T} I_2(t)\leqslant C_2(T-\pi)\|\nu_\cdot^n-\nu_\cdot^{n-1}\|, \tag{4.29}$$

where the constant C_2 only depends on m, λ_b and λ_σ.

Together, equations (4.23), (4.24), (4.28) and (4.29) imply

$$\begin{aligned}&\|\Phi(\nu_\cdot^n)-\Phi(\nu_\cdot^{n-1})\|_T\\&\qquad\leqslant C_1(T-\pi)\|\nu^{n+1}-\nu^n\|_T+C_2(T-\pi)\|\nu^n-\nu^{n-1}\|_T\\&\qquad= C_1(T-\pi)\|\Phi(\nu_\cdot^n)-\Phi(\nu_\cdot^{n-1})\|_T+C_2(T-\pi)\|\nu^n-\nu^{n-1}\|_T,\end{aligned}$$

for positive constants C_1 and C_2, which in turn implies

$$\|\Phi(\nu_\cdot^n)-\Phi(\nu_\cdot^{n-1})\|_T\leqslant C_T\|\nu_\cdot^n-\nu_\cdot^{n-1}\|_T, \tag{4.30}$$

where the constant $C_T > 0$ depends on T. Then, for T small enough the mapping Φ is a contraction. Then, the Banach Fixed Point Theorem implies that it has a unique fixed point $\nu_\cdot^\infty$, which as we mentioned guarantees the uniqueness of the solution to the nonlinear SDE and the corresponding nonlinear martingale problem. Moreover, (4.30) can be written as

$$\|\nu_\cdot^{n+1}-\nu_\cdot^n\|_T\leqslant C_T\|\nu_\cdot^n-\nu_\cdot^{n-1}\|_T,$$

which implies that the sequence $\nu^n_\cdot$ is a Cauchy sequence, and as a result we obtain the following convergence

$$\nu^n_\cdot \to \nu^\infty_\cdot,$$

in the space $C_\xi([0,T], \mathcal{P}(\mathbb{R}^{m+1}))$ equipped with the metric $\|\cdot\|_T$. This, in particular, implies that any convergent subsequence converges to the same limit $\nu^\infty_\cdot$.

Remark 4.2. Observe that the sequence of pairs $(X^{(n)}, X^{(n-1)})_{n\geqslant 1}$ is also tight. This is due to the fact that on products of separable metric tightness follow from tightness in each component (see e.g. Lemma 3.2 in Whitt, 2007). Thus, Prohorov's Theorem implies that the sequence of pairs contains a weakly convergent subsequence[2] $(X^{(n_k)}, X^{(n_k-1)})_{k\geqslant 1}$. Moreover, once again, by the so-called Skorohod representation theorem (see e.g. Theorem 3.1.8 in Ethier and Kurtz, 1986), we may assume that the processes in the subsequece as well as the limiting process $(X^\infty, \tilde{X}^\infty)$ are defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Besides, using the same argument in Remark 4.1 we obtain that for every finite set $\{t_0, t_1 \cdots, t_N\} \subset [0,\infty]$ we have

$$\lim_{k\to\infty} \mathbb{E}[f(X^{(n_k)}_{t_0}, X^{(n_k)}_{t_1}, \cdots, X^{(n_k)}_{t_N})] = \mathbb{E}[f(X^\infty_{t_0}, X^\infty_{t_1}, \cdots, X^\infty_{t_N})], \qquad \text{a.s.}, \tag{4.31}$$

for every f bounded, and in particular, we have

$$\lim_{k\to\infty} \mathbb{E}[f(X^{(n_k)}_t)] = \mathbb{E}[f(X^\infty_t)], \qquad \text{a.s.}, \tag{4.32}$$

for any $f \in C_b(\mathbb{R})$ and $t \geqslant 0$.

Moreover, the processes X^∞ and $\tilde{X}^\infty$ are "the same" in the sense that

$$\mathcal{L}(X^\infty_{\Pi_\cdot}) = \mathcal{L}(\tilde{X}^\infty_{\Pi_\cdot}), \tag{4.33}$$

since

$$\mathcal{L}(X^{(n_k)}_{\Pi_\cdot}) = \nu^{n_k}_\cdot \to \nu^\infty_\cdot,$$

[2]Notice that we are using the same indices n_k just to simplify notation.

and

$$\mathcal{L}(X^{(n_k-1)}_{\Pi_\cdot}) = \nu^{n_k-1}_\cdot \to \nu^\infty_\cdot,$$

thanks to the previous contraction argument.

Finally, the next result provides **Step 4** and finalises the proof.

Lemma 4.6. *The limit process* X^∞ *solves the nonlinear martingale problem corresponding to the triplet* $(\Pi, A[\mu], X^0)$.

Proof. First notice that, due to the way that the sequence of processes (and the sequence of curves that was constructed simultaneously) was defined in Step 1, for every $(n_k)_{k\geqslant 1}$, the process $X^{(n_k)}$ is the unique solution to the martingale problem corresponding to

$$A[\nu^{n_k-1}_s]f(X^{(n_k)}_s) = \frac{1}{2}\sigma^2(X^{(n_k)}_s, \nu^{n_k-1}_s)\frac{d^2}{dx^2}f(X^{(n_k)}_s) + b(X^{(n_k)}_s, \nu^{n_k-1}_s)\frac{d}{dx}f(X^{(n_k)}_s),$$

and

$$\nu^{n-1}_s = \mathcal{L}(X^{(n_k-1)}_{\Pi_s}), \qquad k \geqslant 1.$$

That is, for every $f \in C^2_b(\mathbb{R})$ the process

$$M^{n_k}_t = f(X^{(n_k)}_t) - f(X^{(n_k)}_\pi) - \int_\pi^t A[\nu^{n_k-1}_s]f(X^{(n_k)}_s)ds, \qquad t \geqslant \pi,$$

is a $\mathcal{F}^{(n_k)}$-martingale. Then, we have

$$\mathbb{E}[\eta_{t_1,\cdots,t_N,t_{N+1}}(X^{(n_k)}, \nu^{n_k-1}_\cdot)] = 0 \tag{4.34}$$

where

$$
\begin{aligned}
&\eta_{t_1,\cdots,t_{N+1}}(X^{(n_k)}, \nu_{\cdot}^{n_k-1}) \\
&\quad = \eta_{t_1,\cdots,t_{N+1}}(X^{(n_k)}, \mathcal{L}(X_{\Pi}^{(n_k-1)})) \\
&\quad = \left(f(X_{t_{N+1}}^{(n_k)}) - f(X_{t_N}^{(n_k)}) - \int_{t_N}^{t_{N+1}} A[\mathcal{L}(X_{\Pi_s}^{(n_k-1)})] f(X_s^{(n_k)}) ds \right) \prod_{i=1}^{N} h_i(X_{t_i}^{(n_k)}),
\end{aligned}
$$

whenever $\pi \leqslant t_1 \leqslant \cdots t_N \leqslant t_{N+1}, f \in C_b^2(\mathbb{R}), h_1, \cdots, h_N \in C_b(\mathbb{R})$.

Hence, the ideal candidate to solve the non-linear martingale problem is precisely the limit process X^{∞}. So, we would like to show that

$$
\mathbb{E}[\eta_{t_1,\cdots,t_N,t_{N+1}}(X^{\infty}, \mathcal{L}(X_{\Pi}^{\infty}))] = 0, \tag{4.35}
$$

where

$$
\eta_{t_1,\cdots,t_N,t_{N+1}}(X^{\infty}) = \left(f(X_{t_{N+1}}^{\infty}) - f(X_{t_N}^{\infty}) - \int_{t_N}^{t_{N+1}} A[\mathcal{L}(X_{\Pi_s}^{\infty})] f(X_s^{\infty}) ds \right) \prod_{i=1}^{N} h_i(X_{t_i}^{\infty}), \tag{4.36}
$$

and $\pi \leqslant t_1 < t_2 < \cdots < t_{N+1}$, $f \in C_b^2(\mathbb{R})$, and $h_1, \cdots, h_N \in C_b(\mathbb{R})$. To achieve this, first notice that (4.34) can be written as

$$
\begin{aligned}
0 &= \mathbb{E}\left[\left(f(X_{t_{N+1}}^{(n_k)}) - f(X_{t_N}^{(n_k)}) - \int_{t_N}^{t_{N+1}} A[\nu_s^{n_k-1}] f(X_s^{(n_k)}) ds \right) \prod_{i=1}^{N} h_i(X_{t_i}^{(n_k)}) \right] \\
&= \mathbb{E}\left[f(X_{t_{N+1}}^{(n_k)}) \prod_{i=1}^{N} h_i(X_{t_i}^{(n_k)}) \right] - \mathbb{E}\left[f(X_{t_N}^{(n_k)}) \prod_{i=1}^{N} h_i(X_{t_i}^{(n_k)}) \right] \\
&\quad - \mathbb{E}\left[\left(\int_{t_N}^{t_{N+1}} A[\nu_s^{n_k-1}] f(X_s^{(n_k)}) ds \right) \prod_{i=1}^{N} h_i(X_{t_i}^{(n_k)}) \right].
\end{aligned}
$$

Passing to the limit as k goes to infinity in the last equality, and taking into account equations (4.31) to (4.33), we obtain

$$
\begin{aligned}
0 = \lim_{k\to\infty} \Bigg\{ & \mathbb{E}\left[f(X^{(n_k)}_{t_{N+1}}) \prod_{i=1}^{N} h_i(X^{(n_k)}_{t_i}) \right] - \mathbb{E}\left[f(X^{(n_k)}_{t_N}) \prod_{i=1}^{N} h_i(X^{(n_k)}_{t_i}) \right] \\
& - \mathbb{E}\left[\left(\int_{t_N}^{t_{N+1}} A[\mathcal{L}(X^{(n_k-1)}_{\Pi_s})] f(X^{(n_k)}_s) ds \right) \prod_{i=1}^{N} h_i(X^{(n_k)}_{t_i}) \right] \Bigg\} \\
= \mathbb{E} & \left[f(X^{\infty}_{t_{N+1}}) \prod_{i=1}^{N} h_i(X^{\infty}_{t_i}) \right] - \mathbb{E}\left[f(X^{\infty}_{t_N}) \prod_{i=1}^{N} h_i(X^{\infty}_{t_i}) \right] \\
& - \lim_{k\to\infty} \left\{ \mathbb{E}\left[\left(\int_{t_N}^{t_{N+1}} A[\mathcal{L}(X^{(n_k-1)}_{\Pi_s})] f(X^{(n_k)}_s) ds \right) \prod_{i=1}^{N} h_i(X^{(n_k)}_{t_i}) \right] \right\} \\
= \mathbb{E} & \left[f(X^{\infty}_{t_{N+1}}) \prod_{i=1}^{N} h_i(X^{\infty}_{t_i}) \right] - \mathbb{E}\left[f(X^{\infty}_{t_N}) \prod_{i=1}^{N} h_i(X^{\infty}_{t_i}) \right] \\
& - \left\{ \mathbb{E}\left[\left(\int_{t_N}^{t_{N+1}} A[\mathcal{L}(X^{\infty}_{\Pi_s})] f(X^{\infty}_s) ds \right) \prod_{i=1}^{N} h_i(X^{\infty}_{t_i}) \right] \right\}
\end{aligned}
$$

thanks to the Continuous Mapping Theorem; and this implies that

$$
\mathbb{E}[\eta_{t_1,\cdots,t_m,t_{m+1}}(X^{\infty}, \mathcal{L}(X^{\infty}_{\Pi}))] = 0. \tag{4.37}
$$

Thus, X^{∞} solves the nonlinear martingale problem as required. □

5

A Class of Nonlinear Diffusions with Unbounded Coefficients

> *The source of all good mathematics is the special case, the concrete example. It is frequent in mathematics that every instance of a concept of seemingly great generality is in essence the same as a small and concrete special case.*
>
> — **Paul Halmos**
>
> I want to be a Mathematician (1985)

IN this chapter, we study a class of stochastic processes described by non-linear — in the sense of McKean — stochastic differential equations (SDEs). More precisely, we consider nonlinear SDEs of the form

$$\begin{cases} dX_t = \left(\int \beta(x,u)\mu_t(du)\right) dt + \sqrt{2}dW_t, \quad \mu_t = \mathcal{L}(X_t), \quad \forall t \geqslant 0, \\ X_0 \quad \text{given}, \end{cases}$$

where W denotes a standard Brownian motion defined on some probability space, X_0 is an independent random variable, β is a real continuous function, and μ_t denotes the marginal distribution of the process X at the time t. We investigate different sets of conditions on β which guarantee both existence and uniqueness of solutions to such equations.

5.1 Introduction

The main goal of this chapter is to investigate the existence and uniqueness of solutions to nonlinear SDEs given by

$$\begin{cases} dX_t = b(X_t, \mu_t)dt + \sqrt{2}dW_t, \qquad \mu_t = \mathcal{L}(X_t), \quad \forall t \geqslant 0, \\ X_0 \quad \text{given}, \end{cases} \tag{5.1}$$

where the the drift coefficient is of the form

$$b(x, \mu_t) = \int \beta(x, u)\mu_t(du),$$

with $\beta : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ being a locally bounded measurable function. Here, $W = \{W_t : t \geqslant 0\}$ is a standard d−dimensional Brownian motion defined on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$ and X_0 is a $\mathcal{F}_0$-measurable, $\mathbb{R}^d$-valued, random variable independent of the Brownian motion.

In the literature, the following nonlinear SDE:

$$\begin{cases} dX_t = \beta * \mu_t(X_t)dt + \sqrt{2}dW_t \quad \text{where} \quad \mu_t = \mathcal{L}(X_t), \; t \geqslant 0, \\ X_0 \;\; \text{given}, \end{cases} \tag{5.2}$$

has been extensively studied. Here, the symbol $*$ denotes the convolution operator, i.e.,

$$\beta * \mu(x) = \int \beta(x - u)\mu_t(du),$$

where $\beta : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ is a given measurable function, and both X_0 and $W = \{W_t : t \geqslant 0\}$ are as above. Assuming that there exists a solution to equation (5.2), Itô's formula implies that the family of probability measures $\{\mu_t = \mathcal{L}(X_t), t \geqslant 0\}$ satisfies the weak equation

$$\begin{cases} \dfrac{d}{dt}(f, \mu_t) = (L[\mu_t]f, \mu_t), \quad \forall f \in C_c^2(\mathbb{R}^d) \\ \mu_0 = \xi, \end{cases} \tag{5.3}$$

where

$$(f, \mu) = \int_{\mathbb{R}^d} f(x)\mu(dx),$$

and

$$L[\mu_t]f(x) = \sum_{i=1}^d \frac{\partial^2 f(x)}{\partial x_i^2} + \sum_{i=1}^d \left(\int \beta_i(x-u)\mu_t(du) \right) \frac{\partial f(x)}{\partial x_i}, \qquad t \geqslant 0.$$

Moreover, under suitable conditions on β, the measures $\{\mu_t : t \geqslant 0\}$ have smooth densities $\{u(t, \cdot) : t \geqslant 0\}$ which, by partial integration, form a classical solution to the equation

$$\begin{cases} \dfrac{\partial}{\partial t}u(t,x) = \sum_{i,j=1}^d \dfrac{\partial^2}{\partial x_j \partial x_i}u(t,x) - \sum_{i=1}^d \dfrac{\partial}{\partial x_i}\left[u(t,x)\left(\int \beta_i(x,y)u(t,y)dy\right)\right], \\ u(0,x) = u_0, \end{cases} \tag{5.4}$$

where u_0 is the density of ξ.

Partial Differential Equations (PDEs) of this form appeared as models in mathematical physics. In particular, in 1956 an equation of this form was proposed by Kac (1956) as a stochastic toy model for the Vlasov kinetic equation of plasma. Nowadays equations of this type emerge in different contexts, including communication networks (Dawson, Tang, and Zhao, 2005; McDonald and Reynier, 2006; Graham and Robert, 2009) economics (Dai Pra et al., 2009) and neural networks (Laughton and Coolen, 1995). This connection between nonlinear SDEs of the form (5.2), weak equations of the form (5.3) and nonlinear parabolic equations of the form (5.4) was originally discussed by McKean (1966) under the assumption that β is bounded and globally Lipschitz.

McKean-Vlasov diffusions with bounded and globally Lipschitz coefficients have been studied by several authors and under different assumptions about the specific form of β. For instance, the case when $\beta(x,y)$ depends on x, y just through their difference (then the drift coefficient takes the form of a convolution integral) became of special interest. Stroock and Varadhan (2007) studied the case with β equal to the Dirac measure concentrated on $x-y$. Subsequently, Oelschläger (1985) discussed the case with β equal to the derivative of the Dirac measure at zero. For a general introduction to the study of McKean-Vlasov diffusions with bounded and globally Lipschitz coefficients we refer to Sznitman (1991). He proved existence and uniqueness for the solutions to nonlinear SDEs of the form (5.2) by using a fixed point argument on the set of probability measures defined on the space of continuous functions defined on an interval $[0,T]$ equipped with the Kantorovich-Rubinstein or Wasserstein metric. He also introduced the so-called particle system, i.e.,

$$\begin{cases} dX_t^{i,N} = {}^1\!/_N \sum_{j=1}^N \beta(X_t^{i,N}, X_t^{j,N})dt + \sqrt{2}dW_t^i, \\ \mathcal{L}(X_0^{i,N}) = \xi \quad \text{given}, \end{cases} \tag{5.5}$$

and proved that each $X^{i,N}$ converges to an independent copy of the solution of equation (5.2) as N goes to infinity. This idea provided another way of proving the existence McKean-Vlasov diffusions and also to approximate them. For more information about McKean-Vlasov processes with bounded and globally Lipschitz coefficients we refer to the works of Tanaka and Hitsuda (1981); Méléard (1996) and Bossy and Talay (1997) among others.

In this work, we are interested in the existence of solutions to SDEs of the form (5.2) with unbounded coefficients. As in the case of bounded coefficients there are two main approaches to address this problem: the first one is based on fixed point arguments and the second in a propagations of chaos scheme. On one hand, Benachour et al. (1998) studied SDEs of the form (5.2) in the unidimensional case $(d=1)$ with

$$b(x,\mu) = -\int \beta(x-\mu)\mu_t(du),$$

where $\beta : \mathbb{R} \to \mathbb{R}$ is an odd non-decreasing, locally Lipschitz continuous function with polynomial growth which satisfies

$$\beta(x) - \beta(y) \geqslant \beta_1(x-y) + \beta_0, \qquad \forall x \geqslant y.$$

They proved existence and uniqueness of a solution by using a fixed point argument based on Schauder theorem. Subsequently, Herrmann, Imkeller, and Peithmann (2008) generalised their work to the multidimensional case proving existence of solutions assuming that $\beta : \mathbb{R}^d \to \mathbb{R}^d$ is a locally Lipschitz function with polynomial growth which satisfies the extra condition of being rotationally invariant. These assumptions are different to ours but we recover results similar to theirs using a novel approach —which is also based on a fixed point argument but in different spaces — proposed by Kolokoltsov (Kolokoltsov, 2010 and Kolokoltsov, 2011).

On the other hand, Malrieu (2003) treated SDEs of the form (5.2) when b is of the form

$$b(x, \mu) = -\int \nabla B(x-u)\mu(du),$$

where $B : \mathbb{R}^d \to \mathbb{R}$ is a symmetric, twice differentiable and uniformly convex function and such that its gradient ∇B is a locally Lipschitz and grows polynomially. His objective was to handle, by using a probabilistic approach, the problem of existence of solutions to the partial differential equation

$$\frac{\partial u}{\partial t} = \nabla \cdot [\nabla u + u(\nabla B * u)],$$

where $u(t, x)$ is a time dependent probability measure on $\mathbb{R}^d$. This PDE has a number of physical interpretations. In particular, if $B(x) = \|x\|^3$ and $d = 1$, this equation arises in the modelling of granular media. We refer to Benedetto, Caglioti, and Pulvirenti (1997) for the detailed physical interpretation (see also Benedetto et al., 1998). Subsequently, Cattiaux, Guillin, and Malrieu (2008) generalised the results of Malrieu (2003) to the case when B is no longer uniformly convex. More precisely, they assume that B can be written as the sum of twice differentiable functions, one which is uniformly convex function and the other with compact support.

Let $\beta : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ be a continuous function. The following five assumptions will prove to be sufficient for our results.

Assumption 5.1. (Locally Lipschitz w.r.t. the First Variable) For each $R > 0$ there exists a constant $K_R > 0$ such that

$$|\beta(x,u) - \beta(y,u)| \leqslant K_R|x-y|, \quad \text{whenever } |x| \leqslant R, \ \text{ and } \ |y| \leqslant R, \ \ u \in \mathbb{R}^d.$$

Assumption 5.2. (Polynomial Growth) There exist a constant $K > 0$ and an integer m such that

$$|\beta(x,u)| \leqslant K(1 + |(x,u)|^m), \qquad x, u \in \mathbb{R}^d.$$

Here the notation $|(x,u)|$ refers to the Euclidean norm of the vector (x,u) in the product space $\mathbb{R}^d \times \mathbb{R}^d$.

Assumption 5.3. (Confinement Condition) There exists some $R > 0$ such that

$$x \cdot \beta(x,u) \leqslant 0, \quad \text{whenever } \ |x| \geqslant R, u \in \mathbb{R}^d.$$

Assumption 5.4. (First Condition on Partial Derivatives) All partial derivatives up to the second order of β with respect to u, grow polynomially. That is, there exist constants $K_1, K_2 > 0$ and integers m_1 and m_2 such that

$$\left|\frac{\partial}{\partial u_i}\beta(x,u)\right| \ \leqslant \ K_1(1 + |u|^{m_1}), \quad u \in \mathbb{R}^d, \ \ 1 \leqslant i \leqslant d,$$

and

$$\left|\frac{\partial^2}{\partial u_j \partial u_i}\beta(x,u)\right| \ \leqslant \ K_2(1 + |u|^{m_2}), \quad u \in \mathbb{R}^d, \ \ 1 \leqslant i,j \leqslant d,$$

hold uniformly in x.

Assumption 5.5. (Second Condition on Partial Derivatives) The following two conditions hold:

- For every $x, u \in \mathbb{R}^d$ and $w \in \mathbb{R}^{d\times d}$

$$\operatorname{tr}\left[w^{\mathsf{T}}(D_x\beta(x,u))w\right] < 0.$$

- For every $x, u \in \mathbb{R}^d, w \in \mathbb{R}^{d\times d}$, $z \in \mathbb{R}^{d\times d\times d}$ and $\nu \in \mathcal{P}(\mathbb{R}^d)$

$$\sum_{n,k,i=1}^{d}\left[\sum_{m,l=1}^{d}\left(\int \frac{\partial^2\beta_i(x,u)}{\partial x_l\partial x_m}\nu(du)w_{(ln)}\right)w_{(ln)} + \sum_{m=1}^{d}\left(\int \frac{\partial\beta_i(x,u)}{\partial x_m}\nu(du)\right)z^{(m)}_{(nk)}\right]z^{(i)}_{(nk)} < 0.$$

Our main result is stated as follows.

Theorem 5.1. *Let $\beta : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ be a function which satisfies assumptions 5.1 to 5.5. Then, there exists a unique solution to the nonlinear stochastic differential equation*

$$\begin{cases} dX_t = \left(\int \beta(X_t,u)\mu_t(du)\right)dt + \sqrt{2}dW_t, \qquad \mu_t = \mathcal{L}(X_t); \\ X_0 \quad \textit{given}, \end{cases} \tag{5.6}$$

for every X_0 such that $\mathbb{E}[|X_0|^q] < \infty$ with $q = \max\{m + m_1 + 1, m_2 + 1\}$.

In order to present the proof we require a number of preliminary results which are presented as follows.

5.1.1 Preliminary Results

In this section we present some preliminary material that we require to prove our main result.

First, let us consider the space of probability measures with finite q moments $\mathcal{P}_q$ equipped with the norm induced by its embedding in the dual space of $C_{q,\infty}(\mathbb{R}^d)$. That is,

$$|\nu|_{(C_{q,\infty}(\mathbb{R}^d))^*} = \sup\left\{\left|\int f(u)\nu(du)\right| : f \in C_{q,\infty}(\mathbb{R}^d) \quad \text{and} \quad |f|_q \leqslant 1\right\},$$

for each ν in $\mathcal{P}_q$. Besides, let us set $\xi = \mathcal{L}(X_0)$ and introduce the space of continuous curves of probability measures $C_\xi([0,T],\mathcal{P}_q)$, as defined in Section 2.1, equipped with the norm

$$\|\nu_\cdot\|_{(C_{q,\infty}(\mathbb{R}^d))^*} = \sup_{0\leqslant t\leqslant T} |\nu_t|_{(C_{q,\infty}(\mathbb{R}^d))^*},$$

which defines the distance

$$d_{(C_{q,\infty}(\mathbb{R}^d))^*}(\nu_\cdot,\mu_\cdot) = \sup_{0\leqslant t\leqslant T} |\nu_t - \mu_t|_{(C_{q,\infty}(\mathbb{R}^d))^*}.$$

Remark 5.1. Notice that $C_{q,\infty}(\mathbb{R}^d) \subset C_q(\mathbb{R}^d)$ and consequently $(C_q(\mathbb{R}^d))^* \subset (C_{q,\infty}(\mathbb{R}^d))^*$.

Remark 5.2. Observe that for every $f \in C_{q,\infty}(\mathbb{R}^d)$ with $|f|_q \leqslant 1$, we have

$$\begin{aligned}
\int f(u)\nu_t(du) &= \int \frac{f(u)}{(1+|u|^q)}(1+|u|^q)\nu_t(du) \\
&\leqslant \int |f|_q(1+|u|^q)\nu_t(du) \\
&\leqslant \int (1+|u|^q)\nu_t(du).
\end{aligned}$$

Hence

$$|\nu_t|_{(C_{q,\infty}(\mathbb{R}^d))^*} \leqslant \int (1+|u|^q)\nu_t(du), \quad 0 \leqslant t \leqslant T, \tag{5.7}$$

and of course, taking the supremum over t, we obtain

$$\|\nu_\cdot\|_{(C_{q,\infty}(\mathbb{R}^d))^*} \leqslant \sup_{0\leqslant t\leqslant T} \int (1+|u|^q)\nu_t(du). \tag{5.8}$$

On the other hand, the function $1+|u|^q$ can be approximated by an increasing sequence of function with norm equal to one in the space $C_{q,\infty}(\mathbb{R}^d)$ to obtain the reverse inequality to (5.7), that is

$$\int (1+|u|^q)\nu_t(du) \leqslant |\nu_t|_{(C_{q,\infty}(\mathbb{R}^d))^*},$$

which in turn implies

$$\sup_{0\leqslant t\leqslant T} \int (1+|u|^q)\nu_t(du) \leqslant \|\nu_\cdot\|_{(C_{q,\infty}(\mathbb{R}^d))^*}. \tag{5.9}$$

So, by equations (5.8) and (5.9) we conclude that

$$\|\nu_\cdot\|_{(C_{q,\infty}(\mathbb{R}^d))^*} = \sup_{0\leqslant t\leqslant T} \int (1+|u|^q)\nu_t(du).$$

Lemma 5.2. *Suppose that $\beta : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ is a function satisfying assumptions 5.1 to 5.3 and ξ is a probability measure belonging to $\mathcal{P}_q$ with $q \geqslant m+1$. Then, for every curve of probability measures $\nu_\cdot$ in $C_\xi([0,T],\mathcal{P}_q)$ which satisfies*

$$\|\nu_\cdot\|_{(C_{q,\infty}(\mathbb{R}^d))^*} < \infty;$$

the function $b^\nu : [0,T]\times\mathbb{R}^d \to \mathbb{R}^d$, given by

$$b^\nu(t,x) = \int \beta(x,u)\nu_t(du), \qquad \forall (t,x) \in [0,T]\times\mathbb{R}^d,$$

has the following properties:

1. *b^ν is locally Lipschitz with respect to x;*

2. *b^ν grows polynomially with respect to x;*

3. $\sup_{0\leqslant t\leqslant T}|b^{\nu}(t,0)|<\infty$;

4. There exist a constant $R>0$ *such that*

$$\sup_{0\leqslant t\leqslant T} x\cdot b^{\nu}(t,x)\leqslant 0,$$

whenever $|x|\geqslant R$.

Proof. For (1) observe that

$$\begin{aligned}|b^{\nu}(t,x)-b^{\nu}(t,y)| &= \left|\int\beta(x,u)\nu_t(du)-\int\beta(y,u)\nu_t(du)\right|\\ &\leqslant \int|\beta(x,u)-\beta(y,u)|\,\nu_t(du)\\ &\leqslant K\,|x-y|\,,\end{aligned}$$

whenever $|x|\leqslant R$ and $|y|\leqslant R$ as a consequence of assumption 5.1. Besides, notice that this inequality is uniform in t.

For (2) notice that assumption 5.2 implies that

$$|\beta(x,u)|\leqslant K(1+(|x|+|u|)^m),\qquad \text{for every } x,y\in\mathbb{R}^d,$$

and hence

$$\begin{aligned}|b^{\nu}(t,x)| &\leqslant \int|\beta(x,u)|\,\nu_t(du)\\ &\leqslant K\int 1+(|x|+|u|)^m\nu_t(du)\\ &= K\left[1+\sum_{i=0}^{m}\binom{m}{i}\left(\int|u|^{m-i}\nu_t(du)\right)|x|^i\right]\leqslant K\left(1+\sum_{i=0}^{m}a_i(t)|x|^i\right),\end{aligned}$$

where the quantities

$$a_i(t)=2^m\int|u|^{m-i}\nu_t(du),\qquad i=0,\dots,m,$$

are finite since we are assuming that $\{\nu_\cdot\}$ belongs to $C_\xi([0,\infty),\mathcal{P}_q)$ with $q \geqslant m+1$.

For (3), first notice that

$$\begin{aligned}\sup_{0\leqslant t\leqslant T}|b^\nu(t,0)| &= \sup_{0\leqslant t\leqslant T}\left|\int \beta(0,u)\nu_t(du)\right| \\ &\leqslant \sup_{0\leqslant t\leqslant T}\int|\beta(0,u)|\,\nu_t(du).\end{aligned}$$

Then assumption 5.2 implies that

$$\sup_{0\leqslant t\leqslant T}|b^\nu(t,0)| \leqslant \sup_{0\leqslant t\leqslant T}\int K(1+|u|^m)\nu_t(du). \tag{5.10}$$

Besides, observe that the function $K(1+\|u\|^m)$ belongs to $C_{q,\infty}(\mathbb{R}^d)$ since $q \geqslant m+1$. As a consequence there exists a positive constant κ such that

$$\frac{1}{\kappa}|K(1+|\cdot|^m)|_q = 1.$$

This implies that

$$\int K(1+|u|^m)\nu_t(du) \leqslant \kappa|\nu_t|_{(C_{q,\infty}(\mathbb{R}^d))^*}, \tag{5.11}$$

where the norm on the right hand side is the norm in $\mathcal{P}_q(\mathbb{R}^d)$ induced by its embedding in the dual of $C_{q,\infty}(\mathbb{R}^d)$.

Combining equations (5.10) and (5.11), we obtain

$$\sup_{0\leqslant t\leqslant T}|b^\nu(t,0)| \leqslant \kappa \sup_{0\leqslant t\leqslant T}|\nu_t|_{(C_{q,\infty}(\mathbb{R}^d))^*} = \kappa\|\nu_\cdot\|_{(C_{q,\infty}(\mathbb{R}^d))^*},$$

and this quantity is finite by assumption.

Finally for (4), observe that on one hand

$$x\cdot b^\nu(t,x) \quad = \quad x\cdot\int\beta(x,u)\nu_t(du) = \int x\cdot\beta(x,u)\nu_t(du),$$

by definition, and on the other hand assumption 5.3 implies that

$$\int x \cdot \beta(x,u)\nu_t(du) \leqslant 0,$$

whenever $|x| \geqslant R$. These two facts together imply that (4) holds. $\square$

5.2 Existence and Uniqueness

Lemma 5.3. *Let $\beta : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ be a function which satisfies assumptions 5.1 to 5.3 and ξ a probability measure in $\mathcal{P}_q$ for some $q \geqslant m+1$. Now, consider a curve of probability measures $\nu_.$ belonging to $C_\xi([0,\infty), \mathcal{P}_q)$, and let us introduce the function $b^\nu : [0,\infty) \times \mathbb{R}^d \to \mathbb{R}^d$ given by*

$$b^\nu(t,x) = \int \beta(x,u)\nu_t(du), \qquad (t,x) \in [0,\infty) \times \mathbb{R}^d.$$

Then the SDE given by

$$\begin{cases} dX_t = b^\nu(t, X_t)dt + \sqrt{2}dW_t, \\ X_0 \sim \xi, \end{cases} \tag{5.12}$$

has a unique solution. Moreover, there exists a constant $\lambda > 0$ such that

$$\mathbb{E}[|X_t|^q] \leqslant \mathbb{E}[|X_0|^q]e^{\lambda t}, \tag{5.13}$$

for every $t \geqslant 0$.

Proof. To demonstrate the existence of a unique solution to the SDE in equation (5.12) we are going to show that the corresponding martingale problem is well-posed. That is, for each $(s, x) \in [0, \infty) \times \mathbb{R}^d$ there exists a unique probability measure $\mathbb{P}$ defined on the space of right continuous functions having left limits such that

$$\mathbb{P}(X_s = x) = 1,$$

and the process

$$M_t^f = f(X_t) - \int_s^t L_u f(X_u) du, \qquad t \geqslant s,$$

where

$$L_u f(x) = \sum_{i=1}^d \frac{\partial^2}{\partial x_i^2} f(x) + \sum_{i=1}^d b_i^\gamma(x, u) \frac{\partial}{\partial x_i} f(x), \tag{5.14}$$

is a $\mathbb{P}$-martingale for all functions $f \in C_b^2(\mathbb{R}^d)$. Besides, we will show that the process X is strong Markov. In order to achieve these two aims, we are going to follow the ideas from Theorem 5.2 in Kolokoltsov, 2010 which allows to construct Markov processes for Levy type operators with unbounded coefficients. Therefore, we have to prove that:

(1) The martingale problems for the "normalised" (as defined below) operators are well posed and that the corresponding processes are conservative Feller processes.

(2) There exists a function f_L, in some appropriate weighted space, which acts as a Lyapunov function for the family of operators L_u defined in (5.14). That is, a function f_L such that

$$L_u f_L \leqslant c(f_L + 1),$$

for some positive $c > 0$; and $f_L(x) \to \infty$ as $x \to \infty$. It is worth mentioning that the purpose of the Lyapunov function is to make possible the extension of the well-posedness from the normalised operators to the operators with unbounded coefficients.

Let us start by proving point (1). So, for each integer $m \geqslant 1$, consider the family of "normalised" operators defined as follows

$$L_t^{(m)} f(x) = \sum_{i=1}^{d} \frac{\partial^2}{\partial x_i^2} f(x) + \sum_{i=1}^{d} b^{\nu}_{m(i)}(t,x) \frac{\partial}{\partial x_i} f(x), \qquad t \geqslant 0, \tag{5.15}$$

where

$$b^{\nu}_{m}(t,x) = \chi_p\left(\frac{|b^{\nu}(t,x)|}{m}\right) b^{\nu}(t,x),$$

and $\chi_p : [0,\infty) \to [0,1]$ is a smooth and non-increasing function such that

$$\chi_p(r) = \begin{cases} 1 & \text{if } 0 \leqslant r \leqslant 1, \\ 1/r^p & \text{if } r > 2, \end{cases}$$

for some positive p. Now, notice that for each $m \geqslant 1$ the operators $L_m(t)$ coincide with L_t whenever we have $|b^{\nu}(t,x)| \leqslant m$, by construction. Besides, their diffusion coefficients are all equal to a constant, and their drift coefficients are bounded by construction, and uniformly (with respect to t) Lipschitz continuous with respect to x thanks to assumption 5.1. Then, the martingale problem associated to (5.15) is well-posed for each $m \geqslant 1$. This follows by using, for instance, Theorem 2.1 in Stroock and Varadhan, 1972 (where the martingale problem corresponding to diffusions with bounded and globally Lipschitz continuous was treated) or Theorem 3.1 in Stroock, 1975 (where more general processes with bounded and Lipschitz coefficients were covered). So, for every (s,x) in $[0,\infty) \times \mathbb{R}^d$ there exists a unique probability measure $\mathbb{P}^{(m)}_{s,x}$, on the space of right continuous functions having left limits, such that

$$\mathbb{P}^{(m)}_{s,x}(X_s = x) = 1,$$

and the process

$$M_t^{(m),f} = f(X_t) - \int_0^t L_u^{(m)} f(X_u) du, \qquad t \geqslant 0,$$

is a $\mathbb{P}^{(m)}_{s,x}$-martingale for every $f \in C^2_b(\mathbb{R}^d)$. Moreover, $\mathbb{P}^{(m)}_{s,x}$ is a strong Markov process. In particular, this implies that for very integer m, there exists a unique strong Markov process $X^{(m)}$ which solves the equation

$$\begin{cases} dX^{(m)}_t = \sqrt{2}dW_t + b^{\nu}_m(t, X^{(m)}{}_t)dt, \\ \mathcal{L}(X^{(m)}_0) = \xi. \end{cases}$$

Now we can move to prove part (2). For this, consider the function $f_L : \mathbb{R}^d \to \mathbb{R}$, given by

$$f_L(x) = 1 + |x|^q.$$

It is clear that $f_L(x)$ goes to infinity as $|x| \to \infty$. Besides, a simple calculation gives us

$$L_u f_L(x) = q(q-1)|x|^{q-2} + q|x|^{q-2} b^{\nu}(t, x) \cdot x,$$

and taking into account that β satisfies assumptions 5.1 to 5.3, we can use parts (2), (3) and (4) in Lemma 5.2, to obtain

$$L_u f_L(x) \leqslant q(q-1)|x|^{q-2} + qC|x|^{q-2},$$

where C is a positive constant which only depends on q. This implies that there exists a constant $\lambda > 0$ such that

$$L_u f_L(x) \leqslant \lambda(1 + |x|^q) = \lambda f_L(x), \qquad \text{for all } u \geqslant 0.$$

Hence, $f_L(x)$ can be used as the Lyapunov function for the family of operators L_u that was required in (2). Therefore, Theorem 5.2 in Kolokoltsov, 2011 implies that the martingale problem associated to the family of unbounded operators (5.14) is also well-posed, the corresponding process X is strong Markov, and the inequality (5.13) holds.

$\square$

In particular, Lemma 5.3 guarantees that for every given curve of probability measures $\nu_\cdot$ in $C_\xi([0,T],\mathcal{P}_q)$ with $q \geqslant \max\{m+m_1+1, m_2+1\}$, there exists a unique strong solution to the SDE

$$\begin{cases} dX_t = \int \beta(X_t,u)\nu_t(du)dt + \sqrt{2}dW_t, & t \in [0,T], \\ \mathcal{L}(X_0) = \xi. \end{cases} \tag{5.16}$$

Moreover, the curve of probability measures $\mu_\cdot$, given by

$$\mu_t = \mathcal{L}(X_t), \qquad 0 \leqslant t \leqslant T,$$

belongs to $C_\xi([0,T],\mathcal{P}_q)$ as well. Therefore, we can introduce the mapping $\Phi : C_\xi([0,T];\mathcal{P}_q^N) \to C_\xi([0,T],\mathcal{P}_q^N)$ given by $\Phi(\nu_\cdot) = \mu_\cdot$, where $\mu_\cdot$ is defined as above.

Now, observe that if there exists a solution to the nonlinear SDE given in equation (5.6), then the curve formed by its marginal distributions $\mathcal{L}(X_t)$ for $t \in [0,T]$ constitutes a fixed point for the mapping Φ introduced above. Reciprocally, any fixed point of Φ defines a solution to the nonlinear equation (5.6) on the interval $[0,T]$. This is the key observation behind the proof. So, the idea of the proof consists of proving that Φ is a contraction for a small enough T. As a consequence, Φ has a unique fixed point which defines a unique solution to equation (5.6) on the interval $[0,T]$. To complete the proof we will iterate this construction on the intervals $[T,2T]$, $[2T,3T]$, etc, to obtain a global solution. In order to proceed with this plan we still need a couple of preliminary results. It is worth noticing that this argument only provides a weak solution of the nonlinear SDE. The following two Lemmas will help us to prove that Φ is a contraction.

Lemma 5.4. *Let* $\beta : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ *be a function which satisfies* assumptions 5.1 *to* 5.5*. Let* $X = \{X_t^{s,x}, t \geqslant 0\}$ *be the strong solution to the SDE*

$$X_t^{s,x} = x + \int_s^t b^\nu(u, X_u^{s,x})du + \sqrt{2}\int_s^t dW_u, \qquad t \geqslant s, \tag{5.17}$$

driven by a Brownian motion on a given probability space. Then, the family of linear operators $\{U^{s,t} : t \geqslant s\}$*, given by*

$$U^{s,t}[\nu]f(y) = \mathbb{E}[f(X_t^{s,y})], \qquad 0 \leqslant s \leqslant t,$$

has invariant domain $C^2_{q,m_1,m_2,\infty}(\mathbb{R}^d)$*.*

Proof. The proof is based on the approach proposed in Section 5.3 of Kolokoltsov (2010) which allows to recover the invariant domain of limit processes and semigroups from Levy-type operators with unbounded coefficients. First, let us introduce the processes $Y = \{Y_t : t \geqslant s\}$ and $Z = \{Z_t : t \geqslant s\}$, given by

$$Y_t := \frac{\partial X_t^{s,x}}{\partial x}, \quad \text{i.e.,} \quad Y_t = (Y_{t(ik)}) = \left(\frac{\partial X_t^{s,x}{}_{(i)}}{\partial x_k}\right), \qquad 1 \leqslant i, k \leqslant d,$$

and

$$Z_t := \frac{\partial^2 X_t^{s,x}}{\partial^2 x}, \quad \text{i.e.,} \quad Z_t = (Z_{t(ikn)}) = \left(\frac{\partial^2 X_t^{s,x}{}_{(i)}}{\partial x_n \partial x_k}\right), \qquad 1 \leqslant i, k, n \leqslant d,$$

respectively. Here the derivatives are taken with respect to the initial condition x. The regularity of the trajectories of the process X with respect to the initial point follows from Theorem 4.6.2 in Kolokoltsov, 2011.

Noticing that equation (5.17) can be written by components as

$$X_t^{s,x}{}_{(i)} = x_{(i)} + \int_s^t b_i^\nu(u, X_u^{s,x})du + \sqrt{2}\sum_{j=1}^d \int_s^t \delta_{ij} dW_u^{(j)}, \qquad t \geqslant s, \tag{5.18}$$

for $i = 1, \cdots d$; we can write

$$Y_{t(ik)} = \delta_{ik} + \int_s^t \sum_{m=1}^{d} \frac{\partial b_i^\gamma(u, X_u^{s,x})}{\partial x_m} Y_{u(mk)} du, \qquad t \geqslant s, \quad 1 \leqslant i \leqslant d.$$

for all $1 \leqslant i, k \leqslant d$. More concisely, we can write

$$Y_t = I + \int_s^t b_Y(u, X_u^{s,x}, Y_u) du, \qquad t \geqslant s, \tag{5.19}$$

where

$$b_Y(u, X_u^{s,x}, Y_u)_{(ik)} = \frac{\partial b_i^\gamma}{\partial x}(u, X_u^{s,x}) Y_{u(\cdot k)},$$

and I denotes the identity matrix.

Similarly, we can write Z as

$$Z_t = \int_s^t b_Z(u, X_u^{s,x}, Y_u, Z_u) du, \qquad t \geqslant s, \tag{5.20}$$

where

$$b_Z(u, X_u^{s,x}, Y_u, Z_u)_{(ikn)} = \sum_{m,l=1}^{d} \frac{\partial^2 b_i^\gamma(u, X_u^{s,x})}{\partial x_l \partial x_m} Y_{u(ln)} Y_{u(mk)} + \sum_{m=1}^{d} \frac{\partial b_i^\gamma(u, X_u^{s,x})}{\partial x_m} Z_{u(mkn)},$$

which can be written as

$$b_Z(u, X_u^{s,x}, Y_u, Z_u)_{(ikn)} = \left(Y_u^t \frac{\partial^2 b_i^\gamma(u, X_u^{s,x})}{\partial x^2} W_u \right)_{(nk)} + \left(\frac{\partial b_i^\gamma(u, X_u^{s,x})}{\partial x} \cdot Z_{u(\cdot kn)} \right),$$

for simplicity.

Now, consider the system formed by equations (5.18) and (5.19). Then from Proposition 3.13 in Kolokoltsov (2010) it follows that the solutions are well defined and can be obtained via the Ito–Euler approximation scheme and the solutions form a Feller process whose corresponding family of generators is given by

$$L_t^{X,Y} f(x,y) = \sum_{i=1}^{d} \frac{\partial^2}{\partial x_i^2} f(x,y) + \sum_{i=1}^{d} b_i^{\gamma}(t,x) \frac{\partial}{\partial x_i} f(x,y) + \sum_{m,i,k=1}^{d} \frac{\partial b_i^{\gamma}(t,x)}{\partial x_m} y_{mk} \frac{\partial f(x,y)}{\partial y_{ik}}, \tag{5.21}$$

for every smooth $f : \mathbb{R}^d \times \mathbb{R}^{d\times d} \to \mathbb{R}$. So, we can calculate explicitly the effect of the operator $L_t^{X,Y}$ on

$$f_Y(y) = |y|^2 := \sum_{i,k=1}^{d} y_{ik}^2.$$

obtaining

$$\begin{aligned} L_t^{X,Y}(f_Y(y)) &= \sum_{i,m,k=1}^{d} \frac{\partial b_i^{\gamma}(t,x)}{\partial x_m} y_{mk} \frac{\partial (|y|^2)}{\partial y_{ik}} \\ &= 2 \sum_{i,m,k=1}^{d} \frac{\partial b_i^{\gamma}(t,x)}{\partial x_m} y_{mk} y_{ik}. \end{aligned}$$

Then, assumption 5.5 implies that

$$L_t^{X,Y}(f_Y(y)) \leqslant 0. \tag{5.22}$$

Similarly, consider the system formed by equations (5.18) to (5.20). Then the solutions for this system form a Feller process whose corresponding family of generators is given by

$$\begin{aligned} L_t^{X,Y,Z} f(X,Y,Z) = {} & \sum_{i=1}^{d} \frac{\partial^2}{\partial x_i^2} f(X,Y,Z) + \sum_{i=1}^{d} b_i^{\gamma}(t,X) \frac{\partial}{\partial x_i} f(X,Y,Z) \\ & + \sum_{m,i,k=1}^{d} \frac{\partial b_i^{\gamma}(t,X)}{\partial x_m} y_{mk} \frac{\partial f(X,Y,Z)}{\partial y_{ik}} \\ & + \sum_{n,k,i=1}^{d} \left[\sum_{m,l=1}^{d} \left(\frac{\partial^2 b_i^{\gamma}(t,X)}{\partial x_l \partial x_m} y_{(ln)} \right) y_{(mk)} \right. \\ & \left. + \sum_{m=1}^{d} \left(\frac{\partial b_i^{\gamma}(t,X)}{\partial x_m} z_{(mkn)} \right) \right] \frac{\partial f(X,Y,Z)}{\partial z_{ikn}}. \end{aligned} \tag{5.23}$$

So, we can calculate explicitly the effect of $L_t^{X,Y,Z}$ over the function

$$f_Z(z) = |z|^2 := \sum_{i,k,n=1}^{d} z_{(ikn)}^2,$$

obtaining

$$L_t^{X,Y,Z}(f_Z(z)) = 2\sum_{n,k,i=1}^{d}\left[\sum_{m,l=1}^{d}\left(\frac{\partial^2 b_i^{\gamma}(t,x)}{\partial x_l \partial x_m} y_{(ln)}\right) y_{(mk)} + \sum_{m=1}^{d}\left(\frac{\partial b_i^{\gamma}(t,x)}{\partial x_m} z_{(mkn)}\right)\right] z_{(ikn)}.$$

Moreover, assumption 5.5 implies that

$$L_t^{X,Y,Z}(f_Z(z)) \leqslant 0. \tag{5.24}$$

Now, we are able to prove the result by using Theorem 5.4 in Kolokoltsov (2010). In our case, the diffusion coefficient is constant, the drift coefficient is continuous with polynomial growth, locally bounded, twice differentiable, and globally Lipschitz continuous; and there is no jump coefficient. So, the assumptions of Theorem 3.17 hold locally (as required in Theorem 5.4. Kolokoltsov, 2010). This together with equations (5.22) and (5.24) imply the result by part (iv) of Theorem 5.4 in Kolokoltsov (2010).

□

Lemma 5.5. *Let $\beta : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ be a function which satisfies assumptions 5.1 to 5.4 and set $q = \max\{m + m_1 + 1, m_2 + 1\}$. Then there exists a constant $\kappa > 0$ such that for any two curves of probability measures $\nu_\cdot$ and $\eta_\cdot$ in $C_\xi([0,\infty), \mathcal{P}_r)$ with $r \geqslant q$, the following inequality holds*

$$|L[\nu_t] - L[\eta_t]|_{\mathcal{L}(C^2_{q,m_1,m_2,\infty}(\mathbb{R}^d), C_{q,\infty}(\mathbb{R}^d))} \leqslant \kappa |\nu_t - \eta_t|_{(C^2_{q,m_1,m_2,\infty}(\mathbb{R}^d))^*}, \quad t \geqslant 0, \tag{5.25}$$

where the operators $L[\nu_t]$ *and* $L[\eta_t]$ *are of the form*

$$L[\nu_t] = \sum_{i=1}^{d} \frac{\partial^2}{\partial x_i^2} + \sum_{i=1}^{d} \left(\int \beta(x,y)\nu_t(dy) \right) \frac{\partial}{\partial x_i},$$

and

$$L[\eta_t] = \sum_{i=1}^{d} \frac{\partial^2}{\partial x_i^2} + \sum_{i=1}^{d} \left(\int \beta(x,y)\eta_t(dy) \right) \frac{\partial}{\partial x_i},$$

respectively.

Proof. By definition the left hand side of (5.25) is equal to

$$\sup \left\{ \sup_{x \in \mathbb{R}^d} \frac{|(L[\nu_t] - L[\eta_t])f(x)|}{1+|x|^q} \quad ; \quad f \in C^2_{q,m_1,m_2,\infty}(\mathbb{R}^d), |f|_{C^2_{q,m_1,m_2,\infty}(\mathbb{R}^d)} \leqslant 1 \right\}.$$

For every f in $C^2_{q,m_1,m_2,\infty}(\mathbb{R}^d)$, Cauchy-Schwarz inequality implies that

$$\begin{aligned}
\frac{|(L[\nu_t] - L[\eta_t])f(x)|}{1+|x|^q} &\leqslant \frac{|b^\nu(t,x) - b^\eta(t,x)||\nabla f(x)|}{(1+|x|^q)} \\
&= \frac{|b^\nu(t,x) - b^\eta(t,x)||\nabla f(x)|}{(1+|x|^m)(1+|x|^{m_1})} \frac{(1+|x|^m)(1+|x|^{m_1})}{(1+|x|^q)},
\end{aligned}$$

and using the assumption that $q \geqslant m + m_1 + 1$, we obtain

$$\frac{|(L[\nu_t] - L[\eta_t])f(x)|}{1+|x|^q} \leqslant K \frac{|b^\nu(t,x) - b^\eta(t,x)||\nabla f(x)|}{(1+|x|^m)(1+|x|^{m_1})},$$

for some constant $K > 0$. Therefore,

$$|L[\mu_t] - L[\eta_t]|_{\mathcal{L}(C^2_{q,m_1,m_2,\infty}(\mathbb{R}^d), C_{q,\infty}(\mathbb{R}^d))} \leqslant K \sup_{x \in \mathbb{R}^d} \frac{|b^\nu(t,x) - b^\eta(t,x)|}{1+|x|^m},$$

which can be rewritten as

$$\begin{aligned}
&|L[\mu_t] - L[\eta_t]|_{\mathcal{L}(C^2_{q,m_1,m_2,\infty}(\mathbb{R}^d), C_{q,\infty}(\mathbb{R}^d))} \\
&\qquad \leqslant K \sup \left\{ \left| \int \varphi_x(u)(\nu_t - \eta_t)(du) \right| : x \in \mathbb{R}^d \right\},
\end{aligned}$$

where

$$\varphi_x(u) := \frac{\beta(x,u)}{1+|x|^m}, \qquad \text{for every fixed } \; x \in \mathbb{R}^d.$$

Observing that assumptions 5.2 and 5.4 implies that all the functions $\varphi_x(u)$ belong to $C^2_{q,m_1,m_2,\infty}(\mathbb{R}^d)$, we can conclude that

$$\left|\int \varphi_x(u)(\nu_t-\eta_t)(du)\right| \leqslant \kappa|\nu_t-\eta_t|_{(C^2_{q,m_1,m_2}(\mathbb{R}^d))^*}, \qquad \text{for every } x \in \mathbb{R}^d,$$

where κ is a positive constant, and consequently

$$|L[\nu_t]-L[\eta_t]|_{\mathcal{L}(C^2_{q,m_1,m_2,\infty}(\mathbb{R}^d),C_{q,\infty}(\mathbb{R}^d))} \leqslant (\kappa K)|\nu_t-\eta_t|_{(C^2_{q,m_1,m_2}(\mathbb{R}^d))^*}.$$

$\square$

After these results we are able to present the proof of Theorem 5.1 as follows.

Proof. In order to make notation simpler we will use D and B to denote the spaces of functions $C^2_{q,m_1,m_2,\infty}(\mathbb{R}^d)$ and $C_{q,\infty}(\mathbb{R}^d)$, respectively. The proof is as follows. Let us introduce the mapping Φ on the space $C_\xi([0,T],\mathcal{P}_q)$ equipped with the distance induced by the metric

$$\|\nu_\cdot\|_{D^*} = \sup_{0\leqslant t\leqslant T} |\nu_t|_{D^*},$$

where $|\cdot|_{D^*}$ is the norm of $\mathcal{P}_q$ induced by its embedding in D^*, i.e.

$$|\nu_t|_{D^*} = \sup\left\{\left|\int f(u)\nu_t(du)\right| : f \in D, |f|_D \leqslant 1\right\}.$$

So, define $\Phi : C_\xi([0,T],\mathcal{P}_q) \to C_\xi([0,T],\mathcal{P}_q)$ as follows

$$\Phi(\nu_\cdot) = \mathcal{L}(X^\nu_\cdot),$$

where X^ν is the unique solution to the SDE

$$\begin{cases} dX_t = \int \beta(X_t, u)\nu_t(du)dt + \sqrt{2}dW_t, & t \in [0, T], \\ \mathcal{L}(X_0) = \xi. \end{cases}$$

As we mentioned before, Φ is well defined as a consequence of Lemma 5.3. Besides, the following family of operators

$$U^{s,t}[\nu]f(x) = \mathbb{E}[f(X_t^\nu)|X_s = x], \qquad t \geqslant s \geqslant 0,$$

satisfies the following equations

$$\frac{d}{d\tau}U^{t,\tau}[\nu]f = U^{t,\tau}[\{\nu\}]L[\nu_\tau]f, \qquad \tau \geqslant t \tag{5.26}$$

and

$$\frac{d}{d\tau}U^{\tau,t}[\nu]f = -L[\nu_\tau]U^{\tau,t}[\{\nu\}]f, \qquad \tau \leqslant t \tag{5.27}$$

for all $f \in D$ (which is an invariant domain thanks to Lemma 5.4), where the derivatives exist in the topology of B and are right (or left) side derivatives when appropriate. We refer to Kolokoltsov (2010) for further details about the properties of propagators.

Next, we are going to prove that Φ is a contraction. Then, the Banach fixed point Theorem implies that it admits a unique fixed point. To conclude, we shall prove that this allows us to construct a global solution to the nonlinear equation (5.6).

Let $\nu_\cdot, \eta_\cdot$ be two arbitrary elements of $C_\xi([0, T], \mathcal{P}_q)$. By definition we have

$$\begin{aligned} \|\Phi(\nu_\cdot) - \Phi(\eta_\cdot)\|_{D^*} &= \|\mathcal{L}(X_\cdot^\nu) - \mathcal{L}(X_\cdot^\eta)\|_{D^*} \\ &= \sup_{0\leqslant t\leqslant T} \sup_{|f|_D\leqslant 1} \left| \int f(y)(\mathcal{L}(X_t^\nu) - \mathcal{L}(X_t^\eta))(dy) \right| \\ &= \sup_{0\leqslant t\leqslant T} \sup_{|f|_D\leqslant 1} \left| \mathbb{E}[f(X_t^\nu)] - \mathbb{E}[f(X_t^\eta)] \right|, \end{aligned} \tag{5.28}$$

where X^ν and X^η are the processes corresponding to $\nu.$ and $\eta.$, respectively.

Notice that we can write

$$\begin{aligned}
\mathbb{E}[f(X_t^\nu)] - \mathbb{E}[f(X_t^\eta)] &= \int \mathbb{E}[f(X_t^\nu)|X_0^\nu = x]\xi(dx) - \int \mathbb{E}[f(X_t^\eta)|X_0^\eta = x]\xi(dx) \\
&= \int U^{0,t}[\nu]f(x)\xi(dx) - \int U^{0,t}[\eta]f(x)\xi(dx) \\
&= \int ((U^{0,t}[\nu] - U^{0,t}[\eta])f(x)\xi(dx),
\end{aligned} \tag{5.29}$$

since both process have the same initial distribution. Besides, by using propagator properties (in particular see Theorem 7.3 in Kolokoltsov (2010)) and then rewriting the term as an integral, we obtain

$$\begin{aligned}
(U^{0,t}[\nu] - U^{0,t}[\eta])f &= U^{0,t}[\{\nu\}]U^{t,t}[\eta]f - U^{0,0}[\nu]U^{0,t}[\eta]f \\
&= \int_0^t \frac{d}{d\tau} U^{0,\tau}[\nu]U^{\tau,t}[\eta]f d\tau,
\end{aligned}$$

and by using equations (5.26) and (5.27), we get

$$(U^{0,t}[\nu] - U^{0,t}[\eta])f = \int_0^t U^{0,\tau}[\nu](L[\nu_\tau] - L[\eta_\tau])U^{\tau,t}[\eta]f d\tau. \tag{5.30}$$

Replacing (5.30) in (5.29) and then into (5.28), we obtain

$$\begin{aligned}
\|\Phi(\nu.) - \Phi(\eta.)\|_{D^*} &= \sup_{0 \leqslant t \leqslant T} \sup_{|f|_D \leqslant 1} \left| \int\int_0^t U^{0,\tau}[\nu](L[\nu_\tau] - L[\eta_\tau])U^{\tau,t}[\eta]f d\tau d\xi \right| \\
&\leqslant \sup_{0 \leqslant t \leqslant T} \sup_{|f|_D \leqslant 1} t \left| U^{0,\tau}[\nu](L[\nu_\tau] - L[\eta_\tau])U^{\tau,t}[\eta]f \right|_B \\
&\leqslant T \sup_{0 \leqslant t \leqslant T} \sup_{|f|_D \leqslant 1} \left| U^{0,\tau}[\nu](L[\nu_\tau] - L[\eta_\tau])U^{\tau,t}[\eta]f \right|_B .
\end{aligned}$$

Next, notice that the transformations $U^{\tau,t}[\{\nu\}]$ increases neither the norm in B nor the norm in D since the derivatives of f satisfy the same equation as f. This together with the above expression imply that

$$\begin{aligned}
\|\Phi(\nu_\cdot) - \Phi(\eta_\cdot)\|_{D^*} &\leqslant T \sup_{0\leqslant t\leqslant T} |L[\nu_t] - L[\eta_t]|_{\mathcal{L}(B,D)} \\
&\leqslant T\kappa \sup_{0\leqslant t\leqslant T} |\nu_t - \eta_t|_{D^*} \\
&= T\kappa \|\nu_\cdot - \eta_\cdot\|_{D^*},
\end{aligned}$$

where the penultimate inequality follows by Lemma 5.5. This shows that Φ is a contraction for small enough T. Hence by the Contraction Principle there exists a unique fixed point for this mapping. So, we get a solution to (5.6) on $[0, T]$. To obtain the unique global solution one just has to iterate the construction on the next intervals $[T, 2T]$, $[2T, 3T]$, and the proof is complete. □

6

Two Nonlinear SDEs related to Volatility Models

> *I believe that stochastic methods will transform pure and applied mathematics in the beginning of the third millennium. Probability and statistics will come to be viewed as the natural tools to use in mathematical as well as scientific modelling.*
>
> — **David Mumford**
>
> The Dawning of the Age of Stochasticity (2000)

IN Quantitaive Finance, the Black—Scholes equation is one of the most popular models for pricing vanilla options. The original model, introduced in Black and Scholes (1973) is used to calculate the theoretical price of European put and call options. It assumes the existence of perfect capital markets and that the security prices are log normally distributed or equivalently, the log—returns are normally distributed. More precisely, in the Black-Scholes formulation, asset prices are modelled as a geometric Brownian motion with a constant volatility parameter σ. Through the years the model has been questioned on the basis of the assumption of constant volatility (see e.g. Scott (1987); Rubinstein (1994)) since empirical evidence has showed that the volatility actually varies over the time. Moreover, the idea of stochastic volatility has become popular for derivative pricing and hedging in the last decades. In this chapter we study two kinds of nonlinear SDEs which can be interpreted as stochastic volatility models.

6.1 Introduction

In the classical Black-Scholes model the stock price $S = \{S_t : t \geqslant 0\}$ satisfies

$$\begin{cases} dS_t = rS_t dt + \sigma S_t dW_t, \quad t > 0, \\ S_0 = s > 0 \;\; \text{given}, \end{cases}$$

where $r > 0$ represents the risk free rate, $W = \{W_t : t \geqslant 0\}$ is a standard Brownian motion, and $\sigma > 0$ is a constant which represents the spot volatility of the underlying asset price. This model has been generalised in different ways over the time. For instance, by considering that the volatility is not constant but rather a function depending on both t and S_t, i.e.:

$$\begin{cases} dS_t = rS_t dt + \sigma(t, S_t) S_t dW_t, \quad t > 0, \\ S_0 = s > 0 \;\; \text{given}, \end{cases}$$

where $\sigma : \mathbb{R}^+ \times \mathbb{R} \to \mathbb{R}$, is a positive function which satisfies certain conditions.

In this chapter, we propose two stochastic volatility models under the assumption that the volatility is a function depending on the distribution of the risky asset process in the past time. The purpose is to illustrate that the class of nonlinear stochastic processes that we studied in the previous chapters can arise naturally in simple applications. More precisely, we study nonlinear Stochastic Differential Equations (SDEs) of the following form

$$\begin{cases} dS_t = rS_t dt + \sigma(t, S_t, \mathcal{L}(S_{\Pi_t})) S_t dW_t, \quad t > 0, \\ S_0 = s > 0 \;\; \text{given}, \end{cases} \tag{6.1}$$

where r and s are both positive constants as usual, $\{W_t : t \geqslant 0\}$ is a standard Brownian motion, $\mathcal{L}(S_{\Pi t})$ denotes the distribution of a given delayed vector S_{Π_t} (see Chapter 3 for details on delayed vectors). Thus, our models are determined by SDEs which belong to the class that we studied in Chapter 4.

6.2 Volatility Model I

Our first volatility model is determined by the following nonlinear SDE:

$$\begin{cases} dS_t = rS_t dt + \sigma(t, S_t, \mathcal{L}(S_t, S_{t-\tau}))S_t dW_t, & t > 0, \\ S_0 = s > 0 \ \text{ given}, \end{cases} \tag{6.2}$$

where $r > 0$, and $\sigma(t, S_t, \mathcal{L}(S_t, S_{t-\tau})) = \sqrt{V_t}$, with

$$V_t = \begin{cases} g(t), & 0 \leqslant t < \tau; \\ \frac{1}{\tau} \operatorname{Var}\left[\log \frac{S_t}{S_{t-\tau}}\right], & t \geqslant \tau, \end{cases} \tag{6.3}$$

where τ is a positive constant, and $g : [0, \tau] \mapsto \mathbb{R}^+$ is a continuous function. It is worth noticing that in this case the delay vector is simply

$$S_{\Pi_t} = (S_t, S_{t-\tau}), \qquad t \geqslant \tau,$$

where $\tau > 0$. So, τ plays the role of a delay parameter in the sense that the dynamics of the process at the time t depend on the past via the distribution of the pair $(S_t, S_{t-\tau})$. In the next section we will show that equation (6.2) is related to a particular differential equation of delay type.

6.2.1 Associated Delay Equation

First, let us assume the existence of a strong solution to the SDE (6.2) on some finite interval $[0, T]$ for some $T > \tau$. That is, there is a process $S = \{S_t : 0 \leqslant t \leqslant T\}$ with continuous sample paths and such that

$$\mathbb{P}(S_0 = s) = 1,$$

$$\mathbb{P}\left(\sup_{0 \leqslant t \leqslant T} \int_0^t \{r|S_u| + V_u S_u^2\}\, ds < \infty \right) = 1, \tag{6.4}$$

and

$$S_t = s + \int_0^t rS_u du + \int_0^t \sqrt{V_u} S_u dW_u, \quad 0 \leqslant t \leqslant T, \qquad a.s. \tag{6.5}$$

Next, let us define the process $X = \{X_t : 0 \leqslant t \leqslant T\}$ as follows:

$$X_t = \log S_t, \qquad 0 \leqslant t \leqslant T,$$

and observe that V_t, which is in fact a deterministic function, can be rewritten in terms of X as

$$V_t = \begin{cases} g(t) & 0 \leqslant t < \tau; \\ \dfrac{1}{\tau} \mathrm{Var}\,[X_t - X_{t-\tau}] & \tau \leqslant t \leqslant T. \end{cases} \tag{6.6}$$

Besides, Itô's formula implies that

$$X_t - X_{t-\tau} = r\tau + \int_{t-\tau}^t \sqrt{V_u} dW_u - \frac{1}{2} \int_{t-\tau}^t V_u du, \qquad \tau \leqslant t \leqslant T, \tag{6.7}$$

and consequently

$$\begin{aligned}\mathbb{E}[X_t - X_{t-\tau}] &= \mathbb{E}\left[r\tau + \int_{t-\tau}^{t} \sqrt{V_u}\,dW_u - \frac{1}{2}\int_{t-\tau}^{t} V_u\,du\right] \\ &= r\tau - \frac{1}{2}\int_{t-\tau}^{t} V_u\,du, \end{aligned} \tag{6.8}$$

where the last equation is due to the fact that the process M, defined as the Ito integral

$$M_t = \int_0^t \sqrt{V_u}\,dW_u, \quad 0 \leqslant t \leqslant T,$$

is a martingale. Using equations (6.7) and (6.8) to calculate the second part of (6.6), we obtain

$$\begin{aligned} V_t &= \frac{1}{\tau}\mathrm{Var}\left[X_t - X_{t-\tau}\right] \\ &= \frac{1}{\tau}\mathbb{E}\left[(X_t - X_{t-\tau} - \mathbb{E}[X_t - X_{t-\tau}])^2\right] \\ &= \frac{1}{\tau}\mathbb{E}\left[\left(r\tau + \int_{t-\tau}^{t} \sqrt{V_u}\,dW_u - \frac{1}{2}\int_{t-\tau}^{t} V_u\,du - r\tau + \frac{1}{2}\int_{t-\tau}^{t} V_u\,du\right)^2\right] \\ &= \frac{1}{\tau}\mathbb{E}\left[\left(\int_{t-\tau}^{t} \sqrt{V_u}\,dW_u\right)^2\right] \\ &= \frac{1}{\tau}\mathbb{E}\left[\int_{t-\tau}^{t} V_u\,du\right], \end{aligned}$$

for all $t \in [\tau, T]$, where the last equality is due to the isometry property of the Itô integral. Thus

$$V_t = \frac{1}{\tau}\int_{t-\tau}^{t} V_u\,du, \qquad \tau \leqslant t \leqslant T,$$

and taking the derivative with respect to t in both sides of this expression, we can conclude that V_t satisfies the following delay differential equation

$$V_t' - \frac{1}{\tau}V_t + \frac{1}{\tau}V_{t-\tau} = 0, \qquad \tau \leqslant t \leqslant T. \tag{6.9}$$

This leads us to the following result.

Proposition 6.1. *Suppose that there exists a continuous process* $S = \{S_t : 0 \leqslant t \leqslant T\}$ *defined on some probability space* $(\Omega, \mathcal{F}, \mathbb{P})$*, such that*

$$S_t = s + \int_0^t rS_u du + \int_0^t \sqrt{V_u} S_u dW_u, \quad 0 \leqslant t \leqslant T, \qquad a.s,$$

where

$$V_t = \begin{cases} g(t), & 0 \leqslant t < \tau; \\ \frac{1}{\tau} \operatorname{Var}\left[\log \frac{S_t}{S_{t-\tau}}\right], & \tau \leqslant t \leqslant T, \end{cases}$$

and such that

$$\mathbb{P}\left(\sup_{0 \leqslant t \leqslant T} \int_0^t \{r|S_u| + V_u S_u^2\} ds < \infty\right) = 1, \tag{6.10}$$

Then there is a solution to the following delay differential equation

$$y'(t) - \frac{1}{\tau} y(t) + \frac{1}{\tau} y(t - \tau) = 0, \qquad 0 \leqslant t \leqslant T,$$

with initial condition

$$y(t) = g(t), \qquad 0 \leqslant t < \tau.$$

Proof. The previous discussion shows that the function $V(t) = V_t$ satisfies the delay differential equation and the initial conditon. $\square$

The next proposition guarantees that the reciprocal result holds under certain conditions.

Proposition 6.2. *Consider the ordinary differential equation*

$$y'(t) - \frac{1}{\tau} y(t) + \frac{1}{\tau} y(t - \tau) = 0, \qquad \tau \leqslant t \leqslant T,$$

with initial condition

$$y(t) = g(t), \qquad 0 \leqslant t < \tau,$$

where g *is a continuous function which takes only positive values. Suppose that this differential equation has a positive solution. Then the process* $S = \{S_t : 0 \leqslant t \leqslant T\}$ *given by*

$$S_t = s \exp\left\{r + \int_0^t \sqrt{y(u)} dW_u - \frac{1}{2}\int_0^t y(u) du\right\}, \qquad 0 \leqslant t \leqslant T,$$

satisfies the SDE in equation (6.2) *on* $[0, T]$.

Proof. First of all, the ordinary delay differential equation has a solution by Theorem 3.1 in Bellman and Cooke, 1965. Let $\tilde{y}$ such solution, then we have

$$\tilde{y}(t) = g(t), \quad \text{for all} \;\; 0 \leqslant t < \tau,$$

and

$$\tilde{y}(t) = \frac{1}{\tau}\int_{t-\tau}^t \tilde{y}(u) du, \quad \text{for all} \;\; t \geqslant \tau. \tag{6.11}$$

Now, consider the following SDE

$$\begin{cases} d\tilde{S}_t = r\tilde{S}_t dt + \sqrt{\tilde{y}(t)}\tilde{S}_t dW_t, \quad t > 0 \\ \tilde{S}_0 = s > 0 \;\; \text{given.} \end{cases} \tag{6.12}$$

and observe that this equation has a unique solution given by

$$\tilde{S}_t = \tilde{S}_0 \exp\left\{r + \int_0^t \sqrt{\tilde{y}(u)} dW_u - \frac{1}{2}\int_0^t \tilde{y}(u) du\right\}, \qquad t \geqslant 0.$$

We claim that the process $\tilde{S}$ is a solution to the SDE equation (6.2) on $[0, T]$. To prove this, we start by using Itô's formula to obtain

$$\log \tilde{S}_t = \log \tilde{S}_0 + \int_0^t r du + \int_0^t \sqrt{\tilde{y}(u)} dW_u - \frac{1}{2}\int_0^t \tilde{y}(u) du,$$

for all $t \geqslant 0$, which in turn implies that

$$\log \tilde{S}_t - \log \tilde{S}_{t-\tau} = \int_{t-\tau}^{t} r du + \int_{t-\tau}^{t} \sqrt{\tilde{y}(u)} dW_u - \frac{1}{2} \int_{t-\tau}^{t} \tilde{y}(u) du, \tag{6.13}$$

for all $t \geqslant \tau$. Taking the expectation in (6.13), we get

$$\begin{aligned} \mathbb{E}\left[\log \tilde{S}_t - \log \tilde{S}_{t-\tau}\right] &= \int_{t-\tau}^{t} r du + \mathbb{E}\left[\int_{t-\tau}^{t} \sqrt{\tilde{y}(u)} dW_u\right] - \frac{1}{2}\mathbb{E}\left[\int_{t-\tau}^{t} \tilde{y}(u) du\right] \\ &= \int_{t-\tau}^{t} r du - \frac{1}{2} \int_{t-\tau}^{t} \tilde{y}(u) du, \qquad t \geqslant \tau, \end{aligned} \tag{6.14}$$

where the last line follows by using the martingale property of the Itô integral and the fact that $\tilde{y}$ is a deterministic function. Using equations (6.13) and (6.14), the isometry property of Itô integral, and the fact that $\tilde{y}$ is determinist, we obtain

$$\frac{1}{\tau}\mathbb{E}\left[\left(\log \frac{\tilde{S}_t}{\tilde{S}_{t-\tau}} - \mathbb{E}\left[\log \frac{\tilde{S}_t}{\tilde{S}_{t-\tau}}\right]\right)^2\right] = \frac{1}{\tau} \int_{t-\tau}^{t} \tilde{y}(u) du, \qquad t \geqslant \tau. \tag{6.15}$$

Now, notice that equations (6.11) and (6.15) together imply that

$$\tilde{y}(t) = \frac{1}{\tau}\mathbb{E}\left[\left(\log \frac{\tilde{S}_t}{\tilde{S}_{t-\tau}} - \mathbb{E}\left[\log \frac{\tilde{S}_t}{\tilde{S}_{t-\tau}}\right]\right)^2\right], \qquad t \geqslant \tau. \tag{6.16}$$

Hence, the SDE (6.12) can be written as

$$\begin{cases} d\tilde{S}_t = r\tilde{S}_t dt + \sqrt{\tilde{y}(t)}\tilde{S}_t dW_t, & t > 0, \\ \tilde{S}_0 = s > 0 \ \text{ given}, \end{cases}$$

where

$$\tilde{y}(t) = \begin{cases} g(t) & 0 \leqslant t < \tau; \\ \dfrac{1}{\tau} \operatorname{Var}\left[\log S_t - \log S_{t-\tau}\right] & t \geqslant \tau. \end{cases}$$

and the proof is complete. □

In summary, we have found an ordinary differential equation associated with our nonlinear SDE. Such equation is in fact a differential equation of retarded type, or simply called delay differential equation. This kind of equations have been widely studied in several contexts and applications. In particular, an equation of retarded type may represent the behaviour of a system in which the rate of change of the quantity under investigation depends on the past rate and present values of the quantity.

6.2.2 Generalisation

More generally, we can consider

$$\begin{cases} dS_t = rS_t dt + \sigma(t, S_t, \mathcal{L}(S_t, S_{t-\tau}))S_t dW_t, & t > 0, \\ S_0 = s > 0 \ \text{ given}, \end{cases}$$

where $r > 0$, and $\sigma(t, S_t, \mathcal{L}(S_t, S_{t-\tau})) = \sqrt{V_t}$, with

$$V_t = \begin{cases} g(t), & 0 \leqslant t < n\tau; \\ \frac{1}{n\tau} \sum_{i=1}^{n} \mathbb{E}\left[\left(\log \frac{S_{t-(i-1)\tau}}{S_{t-i\tau}} - \mathbb{E} \log \frac{S_{t-(i-1)\tau}}{S_{t-i\tau}}\right)^2\right], & t \geqslant n\tau, \end{cases}$$

where τ is a positive constant, and $g : [0, \tau] \mapsto \mathbb{R}^+$ is a continuous function. where g is a continuous function defined on $[0, n\tau]$.

Using the same arguments as in the simple case, one can see that this SDEs is equivalent (in the same sense as the simple case) to the delay equation

$$Y'_t - \frac{1}{\tau} Y_t + \frac{1}{\tau} Y_{t-\tau} = 0, \qquad t \geqslant n\tau, \tag{6.17}$$

with initial condition

$$Y_t = g(t), \qquad 0 \leqslant t < n\tau.$$

Thus, both the general, and the simple case can be studied in the same way.

6.2.3 Delay Differential Equation

The following results establish the existence of a unique solution to the delay equation.

Theorem 6.3. *If* $g(t)$ *is a continuous function in* $[0, \tau)$*, then there is a unique solution to the delay equation*

$$y'(t) - \frac{1}{\tau} y(t) + \frac{1}{\tau} y(t-\tau) \quad = \quad 0, \qquad t \geqslant \tau \tag{6.18}$$

with initial condition

$$y(t) \quad = \quad g(t), \qquad 0 \leqslant t < \tau.$$

Moreover, such solution is given by

$$y(t) = e^{(t-\tau)/\tau} \left[g(\tau) - \frac{1}{\tau} \int_0^{t-\tau} e^{-s/\tau} y(s) ds \right], \tag{6.19}$$

for every $t \geqslant \tau$*.*

Proof. Existence and uniqueness follow from Theorem 3.1 in Bellman and Cooke, 1965 which treats the following more general ordinary differential equations

$$a_0 y'(t) + b_0 y(t) + b_1 y(t-\tau) = f(t),$$

for some constants a_0, b_0, and b_1 and f a continuous function. In our case $a_0 = 1, b_0 = -\frac{1}{\tau}, c = \frac{1}{\tau}$ and $f(t) = 0$.

To obtain (6.19), notice that multiplying both sides of (6.18) by $\mu(t) = e^{-t/\tau}$, we obtain

$$\mu(t)\frac{d}{dt}y(t) - \frac{1}{\tau}\mu(t)y(t) + \frac{1}{\tau}\mu(t)y(t-\tau) = 0,$$

which is equivalent to

$$\frac{d}{dt}[y(t)\mu(t)] = -\frac{1}{\tau}\mu(t)y(t-\tau),$$

which in turn implies that

$$y(t) \quad = \quad e^{(t-\tau)/\tau}\left[g(\tau) - \frac{1}{\tau}\int_0^{t-\tau} e^{-s/\tau}y(s)ds\right], \qquad t \geqslant \tau.$$

□

Remark 6.1. It is worth mentioning that from equation (6.19) one can notice that $y(t)$ might take negative values. In such case, we cannot use Proposition 6.2 to construct a solution to the nonlinear SDE since the result requires a positive solution.

Next, we investigate the long time behaviour of the solution to equation (6.18). So, let g be a continuous function, and consider the delay equation

$$y'(t) - \frac{1}{\tau}y(t) + \frac{1}{\tau}y(t-\tau) \quad = \quad 0, \qquad t \geqslant \tau, \tag{6.20}$$

with initial condition

$$y(t) \quad = \quad g(t), \qquad 0 \leqslant t < \tau.$$

The existence and uniqueness of the solution follows from Theorem 6.3. Besides, Theorem 6.3 also shows that such solution satisfies the following equality

$$y(t) = e^{(t-\tau)/\tau}\left[g(\tau) - \frac{1}{\tau}\int_0^{t-\tau} e^{-s/\tau}y(s)ds\right], \tag{6.21}$$

for every $t \geqslant \tau$. In order to study the asymptotic behaviour of y, let us introduce the function $h : \mathbb{R} \to \mathbb{R}$, given by

$$h(s) = s - \frac{1}{\tau} + \frac{1}{\tau}e^{-\tau s}. \tag{6.22}$$

This function is called the characteristic function corresponding to the delay differential equation, and the roots of $h(s) = 0$ are called characteristic roots of the delay differential equation. Lemma 3.2 in Bellman and Cooke, 1965 demonstrates that all the characteristics roots lie in the complex plane to the left of some vertical line. That is, there is a constant $c \in \mathbb{R}$ such that all roots satisfy the condition $Re(s) < c$.

Now, the continuity of g allows us to use Theorem 3.6 in Bellman and Cooke, 1965 which implies that the solution y admits the following representation as a contour integral

$$y(t) = \int_{(c)} e^{(t-\tau)s} \frac{p(s)}{h(s)} ds, \qquad t > \tau, \tag{6.23}$$

where the path of integration is any vertical line $\{Re(s) = c\}$ for which c is sufficiently large to guarantee that all the characteristic roots are located at the left side; and

$$p(s) = g(\tau) - \frac{1}{\tau}\int_0^\tau g(r)e^{-sr} dr. \tag{6.24}$$

Thus,

$$y(t) = \sum_{r=1}^{\infty} \text{Residue}\left(\frac{e^{(t-\tau)s}p(s)}{h(s)}, s = s_r\right), \qquad t > \tau.$$

where $\{s_r\}$ is the sequence of zeros of h arranged in order of decreasing real parts (or increasing imaginary parts or absolute values). Moreover, it is well-known that for each zero s_r we have

$$\text{Residue}\left(\frac{e^{(t-\tau)s}p(s)}{h(s)}, s = s_r\right) = e^{ts_r}q_r(s),$$

where q_r is a polynomial of degree less than the multiplicity of s_r. Thus, $y(t)$ can be written as follows

$$y(t) = \sum_{r=1}^{\infty} e^{ts_r} q_r(s), \qquad t > \tau. \tag{6.25}$$

This representation allows us to known the asymptotic behaviour of y by investigating the characteristic roots of $h(s)$. The following result, which is a consequence of Theorem 4.3 in Bellman and Cooke, 1965, formalises this discussion.

Theorem 6.4. *Suppose that the initial condition g is continuous on the interval $[0, \tau]$, and let y be the continuous solution of Theorem 6.3. Let c any number such that no zeros of $h(s)$ lie on the line $\mathrm{Re}(s) = c$. Then, there is a positive number c_1, independent of t and g, such that*

$$\left| y(t) - \sum_{\mathrm{Re}(s_r) > c} e^{ts_r} q_r(s) \right| \leqslant c_1 m_g e^{ct},$$

where the sum is taken over all characteristic roots s_r to the right of the line $\mathrm{Re}(s) = c$,

$$e^{ts_r} q_r(s) = \mathrm{Residue}\left(\frac{e^{(t-\tau)s} p(s)}{h(s)}, s = s_r \right),$$

and

$$m_g = \max_{0 \leqslant t \leqslant \tau} |g(t)|.$$

Now, we are ready to present the result that provides the asymptotic behaviour of the solution to our delay equation.

Theorem 6.5. *If g is a continuous function, then there exists a unique solution of the delay equation*

$$y'(t) - \frac{1}{\tau} y(t) + \frac{1}{\tau} y(t - \tau) \quad = \quad 0, \qquad t \geqslant \tau, \tag{6.26}$$

with initial condition

$$y(t) = g(t), \qquad 0 \leqslant t < \tau.$$

Moreover, there exist constants $M > 0$, *and* $C > 0$ *such that*

$$\left| y(t) - \left[\frac{2p_0}{\tau} t + \left(\frac{2p_1}{\tau} - \frac{4p_0}{\tau} \right) \right] \right| \leqslant M e^{-Ct}, \qquad \textit{for all } \ t \geqslant 0, \tag{6.27}$$

where

$$p_0 = g(\tau) - \frac{1}{\tau} \int_0^\tau g(s) ds, \qquad \textit{and} \qquad p_1 = \int_0^\tau g(s) ds. \tag{6.28}$$

It thus follows that

i) If $p_0 = 0$, *then* y *converges to* $\frac{2p_1}{\tau}$ *as* t *goes to infinity.*

ii) If $p_0 > 0$, *then* y *diverges to infinity as* t *goes to infinity.*

iii) If $p_0 < 0$, *then* y *diverges to minus infinity as* t *goes to infinity.*

Proof. The existence and uniqueness of the solution follows from Theorem 6.3. Besides, Theorem 6.3 also shows that such solution satisfies the following equality

$$y(t) = e^{(t-\tau)/\tau} \left[g(\tau) - \frac{1}{\tau} \int_0^{t-\tau} e^{-s/\tau} y(s) ds \right], \tag{6.29}$$

for every $t \geqslant \tau$.

In order to prove that equation (6.27) holds, we are going to use the fact that the initial condition function g is continuous, by assumption, and then we can use Theorem 6.4. Let c be any negative constant as in Theorem 6.4. Then, there is a positive number c_1, independent of t and g, such that

$$\left| y(t) - \sum_{Re(s_r) > c} e^{t s_r} q_r(s) \right| \leqslant c_1 m_g e^{ct},$$

where the sum is taken over all characteristic roots s_r to the right of the line $\mathrm{Re}(s) = c$ and

$$m_g = \max_{0 \leqslant t \leqslant \tau} |g(t)|.$$

So, we can take $M = c_1 m_g$. Moreover, a solution $y(t)$ is bounded as t goes to infinity, if and only if

$$\sum_{\mathrm{Re}(s_r)>c} e^{s_r t} q_r(t) = \sum_{\mathrm{Re}(s_r)>c} \text{Residue}\left(\frac{e^{(t-\tau)s} p(s)}{h(s)}, s = s_r\right), \tag{6.30}$$

is bounded as t goes to infinity. Now, notice that among the finite number of roots with $\mathrm{Re}(s_r) > c$, some may have negative real parts. For such roots the terms $e^{s_r t} q_r(t)$ in (6.30) tend to zero as t goes to infinity. So, the only terms which we need to worry about are those corresponding to roots with positive or zero real part.

In our case

$$p(s) = g(\tau) - \frac{1}{\tau} \int_0^\tau g(r) e^{-sr} dr, \tag{6.31}$$

and

$$h(s) = s - \frac{1}{\tau} + \frac{1}{\tau} e^{-\tau s}.$$

So, we can use the fact that $s_r = 0$ is the only characteristic root with non-negative real part and has multiplicity two. This implies that

$$\begin{aligned}
\lim_{t \to \infty} y(t) &= \lim_{t \to \infty} \text{Residue}\left(\frac{e^{(t-\tau)s} p(s)}{h(s)}, s = 0\right) \\
&= \lim_{t \to \infty} q_0(t),
\end{aligned}$$

where $q_0(t)$ is a polynomial of degree less than two. Therefore, $y(t)$ is not necessarily bounded as $t \to \infty$. Actually, y would be bounded as t goes to infinity, if and only if, $q_0(t)$ is a polynomial of degree zero, i.e., equal to a constant.

Therefore, we just have to calculate the residue of

$$f(s) := \frac{e^{(t-\tau)s}p(s)}{h(s)},$$

at $s = 0$. In order to do this we will write the series expansion of $f(s)$ around $s = 0$. That is,

$$\begin{aligned} e^{(t-\tau)s} &= \sum_{k=0}^{\infty} \frac{(t-\tau)^k s^k}{k!}, \\ \frac{1}{h(s)} &= \sum_{k=-2}^{\infty} h_n s^k, \end{aligned}$$

and

$$p(s) = \sum_{k=0}^{\infty} \frac{p^{(k)}(0)}{k!} s^k.$$

Then, we multiply these three series to obtain the series of $f(s)$ and the residue is equal to

$$\text{Residue}\,(f(s), s=0) = \frac{2}{3\tau}\left((3t-2\tau)p(0) + 3p'(0)\right).$$

This equality together with equation (6.31) lead to the result. □

Next we present some examples with different initial conditions to illustrate the different behaviour of the solution to the delay differential equation.

Example 1. (Constant Initial Condition) Let $\sigma_0 > 0$, and consider the delay equation

$$y'(t) - \frac{1}{\tau}y(t) + \frac{1}{\tau}y(t-\tau) = 0, \qquad t \geqslant \tau,$$

with initial condition

$$y(t) \quad = \quad g(t) = \sigma_0, \qquad 0 \leqslant t < \tau.$$

Clearly $g(t)$ is a continuous function in $[0, \tau)$, then Theorem 6.3 implies that there is a unique solution to the delay equation. Moreover,

$$y(t) = \begin{cases} \sigma_0^2, & 0 \leqslant t < \tau; \\ e^{\frac{t-\tau}{\tau}} \left(\sigma_0^2 - \frac{1}{\tau} \int_0^{t-\tau} y(u) e^{-u/\tau} du \right), & t \geqslant \tau. \end{cases}$$

In particular, for $t \in [\tau, 2\tau]$ we obtain

$$\begin{aligned} y(t) &= e^{\frac{t-\tau}{\tau}} \left(\sigma_0^2 - \frac{1}{\tau} \int_0^{t-\tau} \sigma_0^2 e^{-u/\tau} du \right) \\ &= e^{\frac{t-\tau}{\tau}} \left(\sigma_0^2 - \sigma_0^2 (1 - e^{-\frac{(t-\tau)}{\tau}}) \right) = \sigma_0^2, \qquad \tau \leqslant t < 2\tau. \end{aligned}$$

Hence,

$$y(t) = \begin{cases} \sigma_0^2, & 0 \leqslant t < 2\tau; \\ e^{\frac{t-\tau}{\tau}} \left(\sigma_0^2 - \frac{1}{\tau} \int_0^{t-\tau} y(u) e^{-u/\tau} du \right), & t \geqslant 2\tau. \end{cases}$$

In the same way, we can see that $y(t) = \sigma_0^2$ for every $t \in [2\tau, 3\tau]$. Since this process can be repeated as many times as we want, we arrive to the conclusion that $y(t) = \sigma_0^2$ for every $t \geqslant 0$.

Example 2. (Lagrange Polynomial Initial Condition) Let $\sigma_0 > 0$, and consider the delay equation

$$y'(t) - \frac{1}{\tau} y(t) + \frac{1}{\tau} y(t - \tau) \quad = \quad 0, \qquad t \geqslant \tau,$$

with initial condition $g(t)$ given by the Lagrange interpolating polynomial corresponding to five observations $a_0 = 10, a_1 = 11, a_2 = 9, a_3 = 10, a_4 = 20, a_5 = 25$ distributed equidistantly on the interval $[0,\tau]$. That is,

$$\begin{aligned}g(s) = a_0 &+ \frac{(32a_0 s^4)}{(3\tau^4)} - \frac{128a_1 s^4}{3\tau^4} + \frac{64a_2 s^4}{\tau^4} - \frac{128a_3 s^4}{3\tau^4} + \frac{32a_4 s^4}{3\tau^4} - \frac{80a_0 s^3}{3\tau^3}\\ &+ \frac{96a_1 s^3}{\tau^3} - \frac{128a_2 s^3}{\tau^3} + \frac{224a_3 s^3}{3\tau^3} - \frac{16a_4 s^3}{\tau^3} + \frac{70a_0 s^2}{3\tau^2} - \frac{208a_1 s^2}{3\tau^2} + \frac{76a_2 s^2}{\tau^2}\\ &- \frac{112a_3 s^2}{3\tau^2} + \frac{22a_4 s^2}{3\tau^2} - \frac{25a_0 s}{3\tau} + \frac{16a_1 s}{\tau} - \frac{12a_2 s}{\tau} + \frac{16a_3 s}{3\tau} - \frac{a_4 s}{\tau}.\end{aligned}$$

In this case, $g(t)$ is a continuous function in $[0,\tau)$, then Theorem 6.3 implies that there is a unique solution to the delay equation. Moreover,

$$\text{Residue}\left(\frac{e^{ts}p_0(s)}{h(s)}, 0\right) = \frac{At + B\tau}{135\tau},$$

where $A = -21a_0 - 96a_1 - 36a_2 - 96a_3 + 249a_4$ and $B = 14a_0 + 88a_1 + 42a_2 + 136a_3 - 145a_4$. So, Theorem 6.4 implies that

$$y(t) \to \infty,$$

as $t \to \infty$.

6.3 Volatility Model II

In this section we consider a stock price modelled by a Geometric Brownian Motion with random volatility coefficient, that is

$$\begin{cases} dS_t = rS_t dt + Y_t(\mathcal{L}(S_{\leqslant t}))S_t dW_t & t \geqslant 0, \\ S_0 \ \text{ given}, \end{cases} \tag{6.32}$$

where r is a positive constant (which represents the fixed interest rate), $W = \{W_t : t \geqslant 0\}$ is a standard Brownian motion, S_0 is a random variable independent of W, and $Y = \{Y_t : t \geqslant 0\}$ is a stochastic process taking values on a finite state space $\{y_i > 0; i = 1, \cdots, N\}$, such that $Y_t = y_1$ for all $t \leqslant \tau$, and whose intensity matrix Q depends on the distribution of S up to the time t as follows

$$Q_t = q_{ij}(t) = \begin{cases} \mathbb{E}\left[g_{ij}(S_t, S_{t-t_1}, S_{t-t_2}, \cdots, S_{t-t_m})\right] > 0, & \text{if } i \neq j, \\ -\sum_{j \neq i} q_{ij}(t), & \text{if } i = j. \end{cases} \tag{6.33}$$

where each $q_{ij}(t), 1 \leqslant i, j \leqslant N$, represents the transition rate from the state y_i to the state y_j. Besides, we are going to assume that all the functions $g_{ij} : \mathbb{R}^{m+1} \to \mathbb{R}$ are bounded and twice differentiable.

For simplicity, we are going to focus on the case when $N = 2$ and $m = 1$. That is, when Y is a stochastic process taking values in $\{y_1 > 0, y_2 > 0\}$, with $Y_t = y_1$ for all $t \leqslant \tau$, and whose transition matrix is given by

$$Q_t = \begin{pmatrix} -q_{12}(t) & q_{12}(t) \\ q_{21} & -q_{21} \end{pmatrix},$$

where $q_{21} = \lambda > 0$, and $q_{12}(t) = \mathbb{E}[g_{12}(S_t, S_{t-\tau})]$, for a given function g_{12}. Under these assumptions equation (6.32) becomes

$$\begin{cases} dS_t = rS_t dt + y_1 S_t dW_t & 0 \leqslant t \leqslant \tau, \\ dS_t = rS_t dt + Y_t S_t dW_t & t \geqslant \tau, \\ S_0 \ \ \text{given}, \end{cases}$$

where $Y_t = Y_t(\mathcal{L}(S_t, S_{t-\tau}))$. Clearly, we have

$$S_t = S_0 \exp\left\{\left(r - \frac{y_1^2}{2}\right)t + y_1 B_t\right\}, \qquad 0 \leqslant t \leqslant \tau,$$

and

$$\mathbb{E}[|S_t|^n] = \mathbb{E}[|S_0|^n] \exp\left\{\left(nr + n(n-1)\frac{y_1^2}{2}\right)t\right\}, \qquad 0 \leqslant t \leqslant \tau.$$

So, we can write

$$\begin{cases} dS_t = rS_t dt + Y_t(\mathcal{L}(S_t, S_{t-\tau}))S_t dW_t & t \geqslant \tau, \\ S_t = S_0 \exp\left\{\left(r - \frac{y_1^2}{2}\right)t + y_1 B_t\right\} & 0 \leqslant t \leqslant \tau, \\ S_0 \ \text{ given.} \end{cases} \tag{6.34}$$

Thus, this problem can be rigorously formulated by setting a nonlinear martingale problem similar to the one that we studied in Chapter 4. More precisely, in this case we have a family of operators $A[\mu] : C^2(\mathbb{R}) \to C(\mathbb{R} \times \{y_1, y_2\})$, given by

$$A[\mu]f(s, y_i) = \frac{1}{2}(sy_i)^2 \frac{\partial^2}{\partial s^2} f(s, y_i) + rs\frac{\partial}{\partial s} f(s, y_i) + q_{ij}(\xi)(f(s, y_j) - f(s, y_i)), \tag{6.35}$$

$$s > 0, i, j = 1, 2, j \neq i,$$

where $q_{12} = \lambda$ and $q_{21} = \int\int g(s_1, s_2)\mu(ds_1 ds_2)$; which can be written as

$$A[\mu]f(s, y_i) = L[\mu]f(s, y_i) + K[\mu]f(x, y_i),$$

with diffusion part and jump part given by

$$L[\mu]f(s, y_i) = \frac{1}{2}(sy_i)^2 \frac{\partial^2}{\partial s^2} f(s, y_i) + rs\frac{\partial}{\partial s} f(s, y_i), \qquad s > 0, i = 1, 2,$$

and

$$K[\mu]f(s, y_i) = q_{ij}(\mu)(f(s, y_j) - f(s, y_i)), \qquad s > 0, i, j = 1, 2, j \neq i,$$

respectively. So, the nonlinear martingale problem consist of finding a stochastic process $X = (S, Y)$ taking values in $\mathbb{R} \times \{y_1, y_2\}$, defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$, such that

$$f(X_t) - f(X_\tau) - \int_\tau^t A[\mathcal{L}(S_u, S_{u-\tau})]f(X_u)du, \qquad t \geqslant \tau, \tag{6.36}$$

is a martingale with respect to the filtration $\mathcal{F}_t^X$, and

$$S_t = S_0 \exp\left\{\left(r - \frac{y_1^2}{2}\right)t + y_1 B_t\right\} \qquad 0 \leqslant t \leqslant \tau,$$
$$Y_t = y_1 \qquad 0 \leqslant t \leqslant \tau.$$

The difference is that in this case, we have a family of operators with an extra term, corresponding to the jumps part, and both the drift and the diffusion coefficient are unbounded. So, our results cannot be applied as they stand right now. However, this problem suggests that the approach can be extended to cover more general cases.

Final Remarks

The main contributions of this thesis can be summarised as follows.

- In Chapter 3, we established an Itô type formula for processes of the form

$$Z_t = f(X_{t-\pi_m}, \cdots, X_{t-\pi_1}, X_t),$$

where f is a smooth function, X is an Ito-diffusion with bounded and Lipschitz continuous coefficients, and $0 < \pi_1 \leqslant \pi_2 \leqslant \cdots \leqslant \pi_m$ is a collection of delay points. A similar formula was established by Hu, Mohammed, and Yan (2004). However, this reference goes beyond the semimartingale setup of $X = \{X_t : t \geqslant 0\}$, and hence it requires stronger conditions. These conditions, when applied to diffusions X governed by SDEs, would require the diffusion coefficient to be twice Malliavin differentiable. We proved that a modification of their proof is possible if X is a diffusion, leading to weaker conditions allowing for Lipschitz continuous coefficients.

- In Chapter 4, we formulated a nonlinear martingale problem corresponding to the family of operators

$$A[\mu]f(x) = \frac{1}{2}\sigma^2(x, \mu)\frac{d^2}{dx^2}f(x) + b(x, \mu)\frac{d}{dx}f(x),$$

and proved the existence and uniqueness of solutions to this problem under mild assumptions on the coefficients $\sigma(x,\mu) : \mathbb{R} \times \mathcal{P}(\mathbb{R}^{m+1}) \to \mathbb{R}$ and $b(x,\mu) : \mathbb{R} \times \mathcal{P}(\mathbb{R}^{m+1}) \to \mathbb{R}$. Setting the problem in a rigorous way is the first contribution of Chapter 4 since, to the best of my knowledge, this kind of nonlinear martingale problems has not been formally formulated before. Besides, the existence and uniqueness of the solution is provided under bounded and Lipschitz continuity conditions of the coefficients. The proof technique that we proposed can be extended to cover more general cases. This points a possible research line. Indeed, this technique can be extended to analyse the multidimensional case, i.e. when

$$A[\mu]f(x) = \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d}\sigma_{ij}(x,\mu)\frac{\partial^2}{\partial x_i \partial x_j}f(x) + \sum_{i=1}^{d} b_i(x,\mu)\frac{\partial}{\partial x_i}f(x),$$

although this extension requires more involved calculations and those were not covered in this thesis.

- In Chapter 5, we investigated the problem of existence of a class of nonlinear diffusions with unbounded coefficients. More precisely, we provided sufficient conditions to guarantee existence and uniqueness of a weak solution to SDEs of the form

$$\begin{cases} dX_t = \int \beta(X_t,u)\mu_t(du)dt + \sqrt{2}dW_t \quad \text{where} \quad \mu_t = \mathcal{L}(X_t),\ t \geqslant 0, \\ X_0 \ \text{ given.} \end{cases}$$

It is worth pointing out that this problem was approached in a different way than that proposed in Chapter 4 since the drift coefficient is unbounded. However, the contribution consist of studying this kind of equations for the first time. All the previous work on similar SDEs has been done considering $\beta(x,u) = \beta(x-u)$.

- Finally, in Chapter 6, we studied some special cases of nonlinear SDEs inspired by financial mathematical models. In particular, we studied equations of the form

$$\begin{cases} dS_t = rS_t dt + \sqrt{Y_t(\mathcal{L}(S_{\leqslant t}))} S_t dW_t, & t > 0; \\ S_0 \ \text{ given}, \end{cases}$$

where r is a positive constant corresponding to the interest rate, $W = \{W_t : t \geqslant 0\}$ is a standard Brownian motion, and

$$Y_t = \begin{cases} g(t), & 0 \leqslant t < \tau, \\ \frac{1}{\tau}\mathbb{E}\left[\left(\log \frac{S_t}{S_{t-\tau}} - \mathbb{E}\left[\log \frac{S_t}{S_{t-\tau}}\right]\right)^2\right], & t \geqslant \tau, \end{cases}$$

for some fixed constant $\tau > 0$ and a given initial condition $g(t)$. We proved that this problem is equivalent to studying the following ordinary differential equation of retarded type:

$$y'(t) - \frac{1}{\tau}y(t) + \frac{1}{\tau}y(t-\tau) = 0, \qquad t \geqslant \tau,$$

with initial condition $y(t) = g(t)$, for every $0 \leqslant t < \tau$. Using this relationship we proved existence and uniqueness of the solution and investigated its long-time behaviour. Although this kind of SDE is interesting in itself, the long-time behaviour of the solutions showed that such a model might not be very relevant in practice.

Finally, it is worth noting that the examples presented in Chapter 5 and Chapter 6 illustrate that the assumptions on the coefficients for the existence and uniqueness of solutions provided in Chapter 4 are sufficient but not necessary.

Bibliography

Ahn, H. (1997). 'Semimartingale integral representation'. In: *Ann. Probab.* 25.2, pp. 997–1010. URL: http://dx.doi.org/10.1214/aop/1024404427 (cit. on p. 29).

Bellman, R. and K. L. Cooke (1965). *Differential Difference Equations. S. New York/London.* Academic Press (cit. on pp. 134, 137, 139, 140).

Benachour, S., B. Roynette, and P. Vallois (1998). 'Nonlinear self-stabilizing processes II: Convergence to invariant probability'. In: *Stochastic Processes and their Applications* 75.2, pp. 203 –224. URL: http://www.sciencedirect.com/science/article/pii/S0304414998000192 (cit. on p. 4).

Benachour, S., B. Roynette, D. Talay, and P. Vallois (1998). 'Nonlinear self-stabilizing processes I. Existence, invariant probability, propagation of chaos'. In: *Stochastic Processes and their Applications* 75.2, pp. 173 –201. URL: http://www.sciencedirect.com/science/article/pii/S0304414998000180 (cit. on pp. 4, 106).

Benedetto, D., E. Caglioti, and M. Pulvirenti (1997). 'A kinetic equation for granular media'. In: *ESAIM: Mathematical Modelling and Numerical Analysis - Modélisation Mathématique et Analyse Numérique* 31.5, pp. 615–641. URL: http://eudml.org/doc/193851 (cit. on pp. 4, 107).

Benedetto, D., E. Caglioti, J. Carrillo, and M. Pulvirenti (1998). 'A Non-Maxwellian Steady Distribution for One-Dimensional Granular Media'. In: *Journal of Statistical Physics* 91.5-6, pp. 979–990. URL: http://dx.doi.org/10.1023/A%3A1023032000560 (cit. on p. 107).

Billingsley, P. (1999). *Convergence of Probability Measures* (cit. on p. 82).

Black, F. and M. Scholes (1973). 'The Pricing of Options and Corporate Liabilities'. In: *Journal of Political Economy* 81.3, pp. 637–654. eprint: http://dx.doi.org/10.1086/260062. URL: http://dx.doi.org/10.1086/260062 (cit. on p. 128).

Bossy, M. and D. Talay (1997). 'A stochastic particle method for the McKean-Vlasov and the Burgers equation.' In: *Math. Comp.* 66.217, pp. 157–192 (cit. on p. 106).

Carrillo, J. A., R. J. McCann, and C. Villani (2003). 'Kinetic equilibration rates for granular media and related equations: entropy dissipation and mass transportation estimates'. In: *Rev. Mat. Iberoamericana* 19.3, pp. 971–1018. URL: http://projecteuclid.org/euclid.rmi/1077293812.

Cattiaux, P., A. Guillin, and F. Malrieu (2008). 'Probabilistic approach for granular media equations in the non-uniformly convex case'. In: *Probability Theory and Related Fields* 140.1-2, pp. 19–40. URL: http://dx.doi.org/10.1007/s00440-007-0056-3 (cit. on p. 107).

Cont, R. and D.-A. Fournié (2013). 'Functional Itô calculus and stochastic integral representation of martingales'. In: *Ann. Probab.* 41.1, pp. 109–133. URL: http://dx.doi.org/10.1214/11-AOP721 (cit. on pp. 29, 30).

Da Prato, G. and J. Zabczyk (2014). *Stochastic Equations in Infinite Dimensions*. Encyclopedia of Mathematics and its Applications. Cambridge University Press (cit. on p. 29).

Dai Pra, P., W. J. Runggaldier, E. Sartori, and M. Tolotti (2009). 'Large portfolio losses: A dynamic contagion model'. In: *Ann. Appl. Probab.* 19.1, pp. 347–394. URL: http://dx.doi.org/10.1214/08-AAP544 (cit. on p. 105).

Dawson, D. A. and J. Gärtner (1987). 'Large deviations from the Mckean-Vlasov limit for weakly interacting diffusions'. In: *Stochastics* 20.4, pp. 247–308. eprint: http://dx.doi.org/10.1080/17442508708833446. URL: http://dx.doi.org/10.1080/17442508708833446.

Dawson, D., J. Tang, and Y. Zhao (2005). 'Balancing Queues by Mean Field Interaction'. In: *Queueing Systems* 49.3-4, pp. 335–361. URL: http://dx.doi.org/10.1007/s11134-005-6971-z (cit. on p. 105).

Dupire, B. (2009). *Functional Itô Calculus*. Bloomberg Portfolio Research Paper No. 2009-04-FRONTIERS. URL: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1435551 (cit. on pp. 29, 30).

Einstein, A. (1905). 'Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen'. In: *Annalen der Physik* 322.8, pp. 549–560. URL: http://dx.doi.org/10.1002/andp.19053220806 (cit. on p. 16).

Ethier, S. N. and T. G. Kurtz (1986). *Markov Processes : Characterization and Convergence*. Wiley Series in Probability and Mathematical Statistics. New York, Chichester: J. Wiley & Sons (cit. on pp. 88, 89, 99).

Funaki, T. (1985). 'The diffusion approximation of the spatially homogeneous Boltzmann equation'. In: *Duke Math. J.* 52.1, pp. 1–23. URL: http://dx.doi.org/10.1215/S0012-7094-85-05201-9 (cit. on p. 4).

Föllmer, H., P. Protter, and A. N. Shiryayev (1995). 'Quadratic covariation and an extension of Itô's formula'. In: *Bernoulli* 1.1-2, pp. 149–169. URL: http://projecteuclid.org/euclid.bj/1186078365 (cit. on p. 29).

Graham, C. and P. Robert (2009). 'Interacting multi-class transmissions in large stochastic networks'. In: *Ann. Appl. Probab.* 19.6, pp. 2334–2361. URL: http://dx.doi.org/10.1214/09-AAP614 (cit. on p. 105).

Herrmann, S., P. Imkeller, and D. Peithmann (2008). 'Large deviations and a Kramers' type law for self-stabilizing diffusions'. In: *Ann. Appl. Probab.* 18.4, pp. 1379–1423. URL: http://dx.doi.org/10.1214/07-AAP489 (cit. on p. 107).

Hu, Y., S.-E. A. Mohammed, and F. Yan (2004). 'Discrete-time approximations of stochastic delay equations: The Milstein scheme'. In: *Ann. Probab.* 32.1A, pp. 265–314. URL: http://dx.doi.org/10.1214/aop/1078415836 (cit. on pp. 29, 30, 40, 56, 70, 149).

Itô, K. (1944). 'Stochastic integral'. In: *Proc. Imp. Acad.* 20.8, pp. 519–524. URL: http://dx.doi.org/10.3792/pia/1195572786 (cit. on p. 20).

Kac, M. (1956). 'Foundations of Kinetic Theory'. In: *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 3: Contributions to Astronomy and Physics*. Berkeley, Calif.: University of California Press, pp. 171–197. URL: http://projecteuclid.org/euclid.bsmsp/1200502194 (cit. on p. 105).

Karatzas, I. and S. Shreve (1991). *Brownian Motion and Stochastic Calculus*. Graduate Texts in Mathematics. Springer New York (cit. on pp. 8, 31, 77).

Kolokoltsov, V. (2012). 'Nonlinear Lévy and Nonlinear Feller Processes: an Analytic Introduction.' In: *Mathematics and Life Sciences*, pp. 45–70. URL: http://www.degruyter.com/view/books/9783110288537/9783110288537.45/9783110288537.45.xml (cit. on p. 2).

Kolokoltsov, V. (2010). *Nonlinear Markov Processes and Kinetic Equations*. Cambridge Tracts in Mathematics. Cambridge University Press (cit. on pp. x, 4, 31, 73, 107, 115, 119, 120, 122, 125, 126).

– (2011). *Markov Processes, Semigroups and Generators*. De Gruyter Studies in Mathematics. De Gruyter (cit. on pp. 8, 87, 89, 107, 117, 119).

Kunita, H. (1997). *Stochastic Flows and Stochastic Differential Equations*. Cambridge Studies in Advanced Mathematics. Cambridge University Press. URL: https://books.google.co.uk/books?id=_S1RiCosqbMC (cit. on p. 29).

Kunita, H. and S. Watanabe (1967). 'On square integrable martingales'. In: *Nagoya Math. J.* 30, pp. 209–245. URL: http://projecteuclid.org/euclid.nmj/1118796812 (cit. on p. 20).

Laughton, S. and A. Coolen (1995). 'Macroscopic Lyapunov functions for separable stochastic neural networks with detailed balance'. In: *Journal of Statistical Physics* 80.1-2, pp. 375–387. URL: http://dx.doi.org/10.1007/BF02178364 (cit. on p. 105).

Malrieu, F. (2003). 'Convergence to equilibrium for granular media equations and their Euler schemes'. In: *Ann. Appl. Probab.* 13.2, pp. 540–560. URL: http://dx.doi.org/10.1214/aoap/1050689593 (cit. on p. 107).

McDonald, D. R. and J. Reynier (2006). 'Mean field convergence of a model of multiple TCP connections through a buffer implementing RED'. In: *Ann. Appl. Probab.* 16.1, pp. 244–294. URL: http://dx.doi.org/10.1214/105051605000000700 (cit. on p. 105).

McKean, H. P. (1966). 'A Class of Markov Processes Associated with Nonlinear Parabolic Equations'. In: *Proceedings of the National Academy of Sciences of the United States of America* 56.6, pp. 1907–1911. URL: http://www.jstor.org/stable/57643 (cit. on pp. 3, 105).

– (1967). 'Propagation of chaos for a class of nonlinear parabolic equations'. In: *Lecture Series in Differential Equations 2nd ed.* 7, pp. 41–57 (cit. on p. 3).

Métivier, M. (1982). *Semimartingales: A Course on Stochastic Processes*. De Gruyter studies in mathematics. XI. URL: https://books.google.co.uk/books?id=-DUyYSbqK3cC (cit. on p. 29).

Méléard, S. (1996). ‘Asymptotic behaviour of some interacting particle systems; McKean-Vlasov and Boltzmann models’. In: *Probabilistic Models for Nonlinear Partial Differential Equations*. Ed. by D. Talay and L. Tubaro. Vol. 1627. Lecture Notes in Mathematics. Springer Berlin Heidelberg, pp. 42–95. URL: http://dx.doi.org/10.1007/BFb0093177 (cit. on pp. 73, 106).

Nualart, D. (2013). *The Malliavin Calculus and Related Topics*. Probability and Its Applications. Springer New York (cit. on pp. 20, 22, 23, 25–27, 57, 61).

Oelschläger, K. (1985). ‘A law of large numbers for moderately interacting diffusion processes’. In: *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 69.2, pp. 279–322. URL: http://dx.doi.org/10.1007/BF02450284 (cit. on p. 106).

Rubinstein, M. (1994). ‘Implied Binomial Trees’. In: *The Journal of Finance* 49.3, pp. 771–818. URL: http://dx.doi.org/10.1111/j.1540-6261.1994.tb00079.x (cit. on p. 128).

Scott, L. O. (1987). ‘Option Pricing when the Variance Changes Randomly: Theory, Estimation, and an Application’. In: *Journal of Financial and Quantitative Analysis* 22.4, pp. 419–438. URL: https://www.cambridge.org/core/article/option-pricing-when-the-variance-changes-randomly-ory-estimation-and-an-application/D4068ABF9F6527413FD56E66A220BFCA (cit. on p. 128).

Skorokhod, A. V. (1976). ‘On a Generalization of a Stochastic Integral’. In: *Theory of Probability & Its Applications* 20.2, pp. 219–233. eprint: http://dx.doi.org/10.1137/1120030. URL: http://dx.doi.org/10.1137/1120030 (cit. on p. 20).

Stroock, D. W. and S. R. S. Varadhan (2007). *Multidimensional Diffusion Processes*. Classics in Mathematics. Springer Berlin Heidelberg (cit. on pp. 31, 106).

Stroock, D. W. (1975). ‘Diffusion processes associated with Lévy generators’. In: *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 32.3, pp. 209–244. URL: http://dx.doi.org/10.1007/BF00532614 (cit. on pp. 72, 116).

Stroock, D. W. and S. R. S. Varadhan (1969a). ‘Diffusion processes with continuous coefficients, I’. In: *Communications on Pure and Applied Mathematics* 22.3, pp. 345–400. URL: http://dx.doi.org/10.1002/cpa.3160220304 (cit. on p. 72).

– (1969b). ‘Diffusion processes with continuous coefficients, I’. In: *Communications on Pure and Applied Mathematics* 22.3, pp. 345–400. URL: http://dx.doi.org/10.1002/cpa.3160220304.

Stroock, D. W. and S. R. S. Varadhan (1969c). 'Diffusion processes with continuous coefficients, II'. In: *Communications on Pure and Applied Mathematics* 22.4, pp. 479–530. URL: http://dx.doi.org/10.1002/cpa.3160220404 (cit. on p. 72).

– (1972). 'Diffusion processes'. In: *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Volume 3: Probability Theory*. Berkeley, Calif.: University of California Press, pp. 361–368. URL: http://projecteuclid.org/euclid.bsmsp/1200514346 (cit. on pp. 78, 116).

Summers, W. H. (1970). 'Dual Spaces of Weighted Spaces'. In: *Transactions of the American Mathematical Society* 151.1, pp. 323–333. URL: http://www.jstor.org/stable/1995631 (cit. on p. 11).

Sznitman, A.-S. (1991). 'Ecole d' Eté de Probabilités de Saint-Flour XIX -1989'. In: ed. by P.-L. Hennequin. Vol. 1464. Lecture Notes in Mathematics. Springer Berlin Heidelberg, pp. 165–251. URL: http://dx.doi.org/10.1007/BFb0085169 (cit. on pp. 4, 73, 106).

Tanaka, H. and M. Hitsuda (1981). 'Central limit theorem for a simple diffusion model of interacting particles'. In: *Hiroshima Math. J.* 11.2, pp. 415–423. URL: http://projecteuclid.org/euclid.hmj/1206134109 (cit. on p. 106).

Whitt, W. et al. (2007). 'Proofs of the martingale FCLT'. In: *Probab. Surv* 4, pp. 268–302 (cit. on p. 99).

Index